BIG DATA

Discrimination and Data-Driven Decision-Making[1]

Abstract

In the midst of the Black Lives Matter movement and a world-changing pandemic, the Multicultural Media, Telecom and Internet Council (MMTC) will release a series of policy papers discussing technology's role in historic stories of marginalization. This first paper is about *big data*, and how various forms of racism are empowered (and sometimes resisted) by the collection, management, and commercialization of information.

In a surprising variety of social and industry contexts, racist policies and actions are motivated by "objective" data. Even if a piece of data processing technology promises infinite social or financial gains, it should not be called *innovative* until it is accessible by the poor, and the disenfranchised. It is now imperative to invest in tech safety, responsibility, and management, rather than just tech development and marketing.

The examples discussed in this paper are evidence of centuries of under-investment – the consequences of which are evidence that big data should be regulated and developed to avoid perpetuating historic social injustices. The rest of the series will highlight examples of tech-equity issues raised by:

Technology Regulation

Telecommunication and Broadband

Media and the Entertainment Industry

Automation and Artificial Intelligence

[1] Working title.

Contents

Acknowledgements

MMTC would like to thank DeVaughn Jones, MMTC's 2020 Cathy Hughes Fellow, for his leadership and extensive research and writing in getting this series paper to the finish line. This paper could not have been completed without the support of MMTC Staff members and volunteers; notably, Maurita Coley, Ronald Johnson, David Honig, Dorrissa Griffin, Ananda Leeke, Marcella Gadson, Deborah Lathen, Kim Keenan, and former MMTC fellow Ashton Hampton. Special thanks to Maryam Casbarro, CIPP/US of Davis Wright Tremaine, LLP for her expertise and guidance on privacy issues.

Introduction

One of America's least-known civil rights advocates was an immigrant named Albert Einstein.

Twenty years before the John Lewis crossed the bridge at Selma,[2] Einstein taught Black students during the height of his celebrity.[3] Despite the fact that the civil rights movement was still two decades away, he used his celebrity to decry racism in the United States.[4] But sixty years after Selma, John Lewis' name is synonymous with justice – while Albert Einstein remains synonymous with physics.

To be sure, Einstein was a scientific genius; special relativity was his answer to the question, *what is the relationship between space and time?* But social justice is also a science that, like physics, distills to a set of interacting principles. It is unfortunate that science and social justice are siloed throughout the many stories that form American history[5] - but it need not be that way.

---

[2] *See* Brian Van Der Brug, *How Selma's 'Bloody Sunday' Became a Turning Point in the Civil Rights Movement*, HISTORY (Jul. 18, 2020), *available at* https://www.history.com/news/selma-bloody-sunday-attack-civil-rights-movement (last visited Oct. 7, 2020) ("On March 7, 1965 . . . John Lewis led over 600 marchers across the Edmund Pettus Bridge in Selma, Alabama and faced brutal attacks by oncoming state troopers.").

[3] Ken Gewertz, *Albert Einstein, Civil Rights activist*, HARVARD GAZETTE (Apr. 12, 2007), *available at* https://news.harvard.edu/gazette/story/2007/04/albert-einstein-civil-rights-activist/ (last visited Oct. 7, 2020) ("In 1946, [Einstein] traveled to Lincoln University . . . the first school in America to grant college degrees to blacks." There, "Einstein gave a speech in which he called racism 'a disease of white people,' and added, 'I do not intend to be quiet about it.'").

[4] *See id.* ("Einstein continued to support progressive causes through the 1950s, when the pressure of anti-Communist witch hunts made it dangerous to do so.").

[5] *See id.* (revealing that Einstein's visit to Lincoln was "virtually ignored by the mainstream press, which regularly covered Einstein's speeches and activities," and that no major Einstein biographies or archives mentioned the visit, either). In fact, Einstein's "opposition to racism and his relationships with African Americans" account for most of the information missing from "the numerous studies" of his life and work. *Id.*

Background

Technology is as innovative as it is equitably beneficial and accessible. Modern tools like machine learning programs are be "powerful instruments for fighting bias, building progressive movements, and promoting civic engagement."[6] However, *historically* marginalized communities are so-named because they are marginalized throughout history – despite technologies that were commensurately advancing at a more-than-exponential rate.[7] Data-driven discrimination is one modern, technology-driven marginalizing force with an alarming tendency to "perpetuate discrimination and unequal access to opportunity as the use of data expands."[8]

"Big data" is a broad term historically describing large pools of descriptive information, the tools used to derive patterns baked

---

[6] Data for Black Lives, About Us, *available at* https://d4bl.org/about.html (last visited Oct. 7, 2020). *See also* Tyler Danner, *Artificial Intelligence, Machine Learning, and Deep Learning*, AUNALYTICS, *available at* https://www.aunalytics.com/artificial-intelligence-machine-learning-and-deep-learning%E2%81%A0/ (last visited Oct. 7, 2020) (defining machine learning as a "*data-driven* form of artificial intelligence," which is "any rule-based [software] application that simulates human intelligence" by "incorporat[ing] math and statistics in such a way that allows the [AI] to *learn* from data."); Christine Bannan et al., Introduction, *in Automated Intrusion, Systemic Discrimination*, NEW AMERICA (last updated Oct. 7, 2020), *available at* https://www.newamerica.org/oti/reports/automated-intrusion-systemic-discrimination/ (last visited Oct. 7, 2020) ("Automated Intrusion") (explaining that automated decision-making is applied to "decid[e] whether, or for how long, someone should go to jail, whether someone should be considered for an open job, whether someone is likely to succeed at a university, and more.").

[7] *Compare* Staff, *1962: Cordless Phone And Other Tech Innovations Unveiled At Seattle World's Fair*, NPR, *available at* kuow.org/stories/1962-cordless-phone-and-other-tech-innovations-unveiled-seattle-world-s-fair/ (last visited Oct. 7, 2020) ("NASA was ready to launch Telstar [1]" the satellite whose enhanced signal transmission technology enabled the first international television broadcast) *with* Roy Johnson, *What It Was Like to Be an African-American Freshman in 1962*, STANFORD MAGAZINE (Sept. 2017), *available at* https://stanfordmag.org/contents/what-it-was-like-to-be-an-african-american-freshman-in-1962 (last visited Oct. 7, 2020) ("Although it had been eight years since Brown v. Board of Education . . . Ole Miss still admitted only white students.").

[8] *See* Report, White House, Executive Office of the President, *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights* at 6 (May 2016), *available at* https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf (last visited Oct. 7, 2020) (hereinafter, "*Big Data: A Report on Algorithmic Systems*").

into that information, and the actors[9] who leverage both the tools and the information to derive valuable patterns.[10] Data is valuable when there is a lot of it, when it can be easily analyzed, and when it comes from many different sources.[11] It follows that data-related

[9] For example, Amazon is a "big data" company because its primary business activity is collecting and analyzing all kinds of data using whatever is in its algorithmic toolbox. Amazon can let its customers use those tools for a price – or, it can leverage the analysis to make extremely accurate decisions about how to provide products to its users. *See* Sooraj Shah, *Where the money is really made at Amazon*, BBC NEWS (Dec. 13, 2019), *available at* https://www.bbc.com/news/business-50728082 (last visited Oct. 7, 2020) ("[Amazon Web Services] sells data storage and processing for companies that don't [to run] their own IT infrastructure. It's a business known as cloud computing and . . . AWS supplied 70% of Amazon's profits in its most recent quarter."). *See also* Wheeler, *infra* note 14 (describing the biggest internet companies' ability to collect data because the data "fuels the software algorithms that deliver the companies' services.")

[10] Recent research fails to produce a uniform definition of "Big Data" among experts. Surveyed national regulatory definitions place the data itself as the object of the term, *see*, *e.g.*, Fact Sheet, Eur. Comm'n, *The EU Data Protection Reform and Big Data* (Jan. 2016), *available at* https://op.europa.eu/en/publication-detail/-/publication/51fc3ba6-e601-11e7-9749-01aa75ed71a1 (last visited on Oct. 7, 2020); U.S. Nat'l Science Found., Nat'l Institutes of Health, *Core Techniques and Technologies for Advancing Big Data Science & Engineering (BIGDATA)*, *available at* https://www.nsf.gov/pubs/2012/nsf12499/nsf12499.pdf (last visited Oct. 7, 2020), while academics have a variety of definitions that also incorporate the data's use and user. *See* Maddalena Favaretto et al., *What is your definition of Big Data? Researchers' understanding of the phenomenon of the decade*, PLOS ONE (Feb. 25, 2020), *available at* https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0228987 (last visited Oct. 7, 2020) ("Big Data is an interdisciplinary field that requires the connection of different disciplines" and "heterogeneous research skills" to "fully exploit" the benefits of the underlying data).

[11] Foundational scholarship analyzes data according to its volume, velocity, and variety. Variety relates to the diversity of a dataset's sources; velocity refers to the speed which that data can be analyzed; volume refers to the data's depth and breadth. *See generally* Doug Laney, *3D Data Management: Controlling Data Volume, Velocity, and Variety*, META GROUP, File: 949 (Feb. 6, 2001), *available at* https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf (last visited Oct. 7, 2020). *See also* Esteban Ortiz-Ospina, *The rise of social media*, OUR WORLD IN DATA (Sept. 18, 2019), *available at* https://ourworldindata.org/rise-of-social-media (last visited Oct. 7, 2020) ("[T]here are 7.7 billion people in the world, with at least 3.5 billion of us online. This means social media platforms are used by one-in-three people in the world, and more than two-thirds of all internet users."); *IDC Forecasts Revenues for Big Data and Business Analytics Solutions Will Reach $189.1 Billion This Year with Double-Digit Annual Growth Through 2022*, INT'L DATA CORP. (Apr. 4, 2019), *available at* https://www.idc.com/getdoc.jsp?containerId=prUS44998419 (last visited Oct. 7, 2020) (forecasting worldwide big data revenues to reach over a quarter-trillion dollars by 2022); YouTube, *YouTube by the numbers*, *available at*

technologies have become more advanced as the amount of actionable data in the world has increased – machine learning systems, for example, are deployed where the volume of data to be analyzed exceeds human capacity.[12] Big data may be compared to concepts like 'big pharma,'[13] especially in the context of today's technology industry.[14]

Like many of its macroeconomic contemporaries, big data sparks criticism and skepticism.[15] Data-driven discrimination has shown that objectively correct user data is used to justify intentionally racist corporate behavior.[16] In many cases, though, big data simply takes discriminatory patterns from the real-world and ports them into online spaces.[17] Aggrieved users have already begun to gain traction in court because they face discrimination online that

---

https://www.youtube.com/about/press/ (last visited Sept. 10, 2020), (reporting that YouTube has over two billion users and streams over one billion hours of video per day, or about 42 million days' worth of video every 24 hours).

[12] *See*, *e.g.*, YouTube, *YouTube by the numbers*, *supra* note 11.

[13] *See* Definition, Big Pharma, Cambridge Dictionary, *available at* https://dictionary.cambridge.org/us/dictionary/english/big-pharma (last visited Sept. 10, 2020), ("[L]arge pharmaceutical companies[,] companies producing medical drugs[,] especially when these are seen as having a powerful and bad influence.").

[14] *See* Tom Wheeler, *Big Tech and antitrust: Pay attention to the math behind the curtain*, BROOKINGS (Jul. 31, 2020), *available at* https://www.brookings.edu/blog/techtank/2020/07/31/big-tech-and-antitrust-pay-attention-to-the-math-behind-the-curtain/ (last visited Oct. 7, 2020) (describing "the source" of major tech companies' power as "the collection and hoarding of the digital data that fuels the software algorithms that deliver the companies' services. It is the 21st century equivalent of Rockefeller's 20th century monopoly over oil.").

[15] *See, e.g.*, Determination, *Bradley v. Capital One*, Charge Number 570-2018-01036 (EEOC Jul. 2019) (finding Capitol One unlawfully discriminated by advertising jobs on Facebook while limiting the age of people who could see the advertisement) *with Bradley v. T-Mobile*, Case No. 17-cv-07232-BLF, 2019 WL 2358972 *1, *5 (N.D. Cal. 2020) (arguing that companies used Facebook's advertising tools to unlawfully discriminate against older users). *See also* Carroll Doherty et al., *Americans have become much less positive about tech companies' impact on the U.S.*, PEW RESEARCH CENTER (Jul. 29, 2019), *available at* https://www.pewresearch.org/fact-tank/2019/07/29/americans-have-become-much-less-positive-about-tech-companies-impact-on-the-u-s/ (last visited Oct. 7, 2020).

[16] *See* Three Case Studies, *supra* note 20 et seq. and accompanying text.

[17] *See generally* Mark MacCarthy, *Fairness in algorithmic decision-making*, BROOKINGS (Dec. 6, 2019), *available at* https://www.brookings.edu/research/fairness-in-algorithmic-decision-making/ (discussing several theories on how algorithms can create, perpetuate, or amplify underlying biases hidden in the data they process).

reflects discrimination in real-life, but algorithms were doing the discriminating.[18] Even where the underlying data presents no real-world discriminatory pattern to replicate, an algorithm can still be possessed by the conscious or unconscious bias of its creator.[19]

Make no mistake: discrimination did not start with big data, and data-driven decision-making is not inherently discriminatory. Rather, honest criticism is necessary for innovation to happen. Data is being used to justify consequential decisions – and in several consequential contexts, there is evidence that those data-driven decisions are making discrimination worse. In order to ensure *positive* innovation, rather than innovation for innovations' sake, those contexts must be identified, studied, and leveraged to guide future big data development.

---

[18] See Complaint, *Newman v. Google*, No. 5:20-cv-04011 (N.D. Cal., filed Jun. 16, 2020) (alleging YouTube's algorithms target black creators); *Divino Group v. Google*, No. 5:2019cv04749 (N.D. Cal., filed Aug. 13, 2019) (alleging YouTube discriminates against LGBTQ+ creators).

[19] *E.g.*, Heidi Ledford, *Millions of black people affected by racial bias in health-care algorithms: Study reveals rampant racism in decision-making software used by US hospitals — and highlights ways to correct it*, NATURE (Oct. 24, 2019), *available at* https://www.nature.com/articles/d41586-019-03228-6 (last visited Oct. 7, 2020) (describing how an algorithm used by hospitals and insurers "to help manage care for about 200 million people in the United States each year" directly caused Black patients to be less likely to receive personalized care than White patients). *Id.* ("The scientists speculate that this reduced access to care is due to the effects of systemic racism, ranging from distrust of the health-care system to direct racial discrimination by health-care providers."). *Compoare id. with* Introduction *in* Automated Intrusion, *supra* note 1 ("Because [algorithmic] systems rely on large datasets and statistical analyses, their outputs are often perceived as neutral and not affected by biases in the same way as human decision-making.").

Three Case Studies

Throughout the 20th century, data has been used to entrench discriminatory practices purportedly banned by law. Some notorious examples of discriminatory data use involve **market redlining, voting rights, and law enforcement**. The lessons learned from these toxic practices must inform immediate policy changes; this is especially true in the context of the Black Lives Matter movement and the contemporaneous COVID-19 pandemic, both of which have emphasized how the American status quo often perpetuates human suffering as a result of insufficient or unsophisticated data collection and analysis.

Market Redlining

Coined in the 1960s, the term *redlining* derives from the literal use of red lines on maps to delineate "at risk" or dangerous areas where, for example, banks might not want to offer home mortgages. Invariably, the red lines corresponded with the borders of Black neighborhoods.[20] Redlining entered common parlance after more than three decades of racist housing practices, which were initially spearheaded by the Home Owners Loan Corporation (HOLC) in the 1930s.[21]

---

[20] *See, e.g.*, Lisa Rice & Deirdre Swesnik, National Fair Housing Alliance, *Discriminatory Effects of Credit Scoring on Communities of Color*, 46 SUFFOLK U. L. REV. 935, 940 (Jun. 2012), *available at* http://www.nationalfairhousing.org/LinkClick.aspx?fileticket=yg7AvRgwh%2F4%3D&tabid=3917&mid=5418 (last visited Oct. 7, 2020) (hereinafter, "*Discriminatory Effects*"). *See also* RICHARD ROTHSTEIN, THE COLOR OF LAW: A FORGOTTEN HISTORY OF HOW OUR GOVERNMENT SEGREGATED AMERICA 67, 97, 108, 113 (2017) (hereinafter, "COLOR OF LAW"). *See also What Does 'Black Lives Matter' Mean?*, WALL ST. J. (Aug. 11, 2020), *available at* https://www.wsj.com/articles/what-does-black-lives-matter-mean-11597188691 (last visited Oct. 7, 2020) (framing the Black Lives Matter movement as "an effort to recognize that practices like . . . redlining are meant to silence black voices, that black people led the fight for many of the freedoms we enjoy today, and that we hold privileges George Floyd, Breonna Taylor and Trayvon Martin never had."); Emily Badger, *Redlining: Still a Thing*, WASH. POST (May 28, 2015), *available at* https://www.washingtonpost.com/news/wonk/wp/2015/05/28/evidence-that-banks-still-deny-black-borrowers-just-as-they-did-50-years-ago/?utm_term=.ab3f7f3e365f (last visited Oct. 7, 2020) ("Neighborhoods were ranked and color-coded, and the D-rated ones – shunned for the "inharmonious racial groups – were typically outlined in red.").

[21] *See, e.g.*, Amy E. Hillier, *Redlining and the Home Owner's Loan Corporation*, 29 J. OF URB. HIST. 395 (May 2003), *available at* http://repository.upenn.edu/cgi/viewcontent.cgi?article=1002&context=cplan_papers (last visited Oct. 7, 2020) (hereinafter, "*Redlining and the Home Owner's*

Housing redlining is still prominent today, and worse, there is ample evidence that the practice has evolved to infect the technology and communications industries.[22] Redlining is now a central issue motivating the Black Lives Matter movement – a motivation that portends imminent reforms to how companies and governments are allowed to allocate resources according to data collection.[23]

The Federal Housing Administration (FHA) created HOLC to "stimulate the housing market by removing poor performing mortgages from the U.S. housing market."[24] To assist in deciding which mortgages it would refinance, HOLC developed a "discriminatory risk rating system whereby prospective borrowers were favored if their neighborhood was deemed 'new, homogeneous, and in demand in good times and bad.'"[25] Demographic and "real estate risk" surveys, conducted in over 230 cities across the country, were used to collect the data underlying these maps.[26]

HOLC appraisers divided cities into districts rated by risk of default. Areas with even relatively small Black populations were usually given the lowest rating indicating a high risk of default."[27] The data these appraisers collected included "occupation, income, and ethnicity of [each home's] inhabitants and the age, type of

---

Loan Corporation"). *See also* Naa Oyo A. Kwate et al., *Retail Redlining in New York City: Racialized Access to Day-to-Day Retail Resources*, 90 J. URB. HEALTH: BULL. OF THE N. Y. ACAD. OF MED. 632, 633 (2012), *available at* https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3732689/pdf/11524_2012_Article_9725.pdf (last visited Oct. 7, 2020) (describing food, pharmacy, and banking deserts in low-income neighborhoods as symptoms of retail redlining).

[22] *See infra* note 39 et seq.

[23] *See What Does 'Black Lives Matter' Mean?*, *supra* note 20.

[24] Brian Gilmore, *Chances Are: Lessons from the 1962 United States Civil Rights Commission Housing Discrimination Hearings in Washington, D.C. for the Current Foreclosure Crisis,* 3 COLUM. J. RACE & L. 1, 10-11 (2013) (hereinafter, "*Chances Are*"), *available at* https://journals.library.columbia.edu/index.php/cjrl/article/view/2277/1179 (last visited Oct. 7, 2020) (citing Jackson, *Race, Ethnicity, and Real Estate Appraisal*, *infra* note 28, *in* BECKY NICOLAIDES ET AL., THE SUBURB READER at 247,247–48 (Becky M. Nicolaides & Andres Wiese eds., 2006).

[25] *See Discriminatory Effects*, *supra* note 20, at 7.

[26] *See Redlining and the Home Owner's Loan Corporation*, *supra* note 21, at 394-95.

[27] *See* Michael H. Schill & Susan M. Wachter, *The Spatial Bias of Federal Housing Law and Policy: Concentrated Poverty in Urban America,* 143 U. PA. L. REV. 1285, 1309 (1995).

construction, price range, sales demand, and general state of repair of the housing stock."[28] This data was then categorized and plotted on color-coded maps, with areas shaded in red deemed high-risk for lenders.[29]

Racist and long-lasting policies were perpetuated under the guise of conclusions drawn from "data analysis." Redlining cemented the geographic contours of racial segregation in the United States, the topography of which was long influenced by blatantly racist real estate practices and legal frameworks.[30] It created a preference for new suburban developments populated exclusively by Whites, creating the first crack in the urban-suburban racial split that would widen significantly in the 1950s and 1960s.[31]

Moreover, the inability to access mortgages by Blacks meant they were unable to become homeowners at the same rate as Whites, thereby impacting their ability to "use their homes to accumulate wealth" and further broadening the wealth gap between the two groups.[32] Finally, the approach developed by HOLC served as the basis upon which subsequent mortgage policies were based, creating a self-reinforcing cycle of discrimination and segregation that had at least some semblance of objectivity because it was grounded in

---

[28] *See* Kenneth T. Jackson, *Race, Ethnicity, and Real Estate Appraisal: The Home Owners Loan Corporation and the Federal Housing Administration*, 6 J. OF URB. HIST. 419, 422 (Aug. 1980) (hereinafter, "*Race, Ethnicity, and Real Estate Appraisal*").

[29] *See Discriminatory Effects*, *supra* note 20, at 7.

[30] *See, e.g.,* Daniel Aaronson, Daniel Hartley, & Bhash Mazumder, Federal Reserve Bank of Chicago, *The Effects of the 1930s HOLC "Redlining" Maps*, Working Paper 2017-12 (Aug. 2017), *PDF download available at* https://www.chicagofed.org/~/media/publications/working-papers/2017/wp2017-12-pdf (last visited Oct. 7, 2020). *See also* COLOR OF LAW, *supra* note 20, at 67, 97, 108, 113; IRA KATZNELSON, WHEN AFFIRMATIVE ACTION WAS WHITE: AN UNTOLD HISTORY OF RACIAL INEQUALITY IN TWENTIETH CENTURY AMERICA 163 (2005) ("[F]ederally backed [institutions], especially those created by the GI Bill for veterans, used redlining . . . and overt discrimination to make it very difficult, often impossible, for blacks to qualify for mortgages.").

[31] *See, e.g.,* Benjamin Zimmer, *A Deregulatory Framework for Alleviating Concentrated African-American Poverty*, 9 HASTINGS RACE & POVERTY L. J. 555, 562-63 (2012).

[32] *See* Reginald Leamon Robinson, *Poverty, The Underclass, and the Role of Race Consciousness: A New Age Critique of Black Wealth/White Wealth and American Apartheid*, 34 IND. L. REV. 1377, 1395 (2001).

data.[33] These practices, which became widespread by the 1960s, were supposed to have been outlawed by the Fair Housing Act of 1968.

But in 2020, it has proven difficult to undo the damage done by historic redlining – and to stop continued redlining from happening. [34] Neighborhoods remained segregated after the 1960s, which "starved [low-income and minority areas] of affordable capital" and contributed, along with numerous other factors, to the creation of the "ghettos and housing projects that litter our inner-cities."[35] These conditions hastened the rise and fostered popularity of subprime mortgages in these communities, a lending practice that played a major role in precipitating the Great Recession of 2008.[36] In the aftermath of the recession, "the typical black household [lost] 40 percent of non-home-equity wealth,"[37] further reinforcing the unfortunate cycle that began with HOLC's data-driven approach in the 1930s.[38]

---

[33] *Race, Ethnicity, and Real Estate Appraisal*, *supra* note 24, at 422. *See also Chances Are*, *supra* note 24, at 10-11; *Redlining and the Home Owner's Loan Corporation*, *supra* note 21, at 394-95.

[34] *See*, *e.g.*, The Fair Housing Project of Greater Boston, 1968-Present: Housing Discrimination, *available at* http://www.bostonfairhousing.org/timeline/1968-Housing-Discrimination-Today.html (last visited Oct. 7, 2020).

[35] Many of the discriminatory practices originally legitimatized with data stemming from supposedly objective surveys and risk assessments remain in use today. *See* Jamelle Bouie, *How We Built the Ghettoes*, DAILY BEAST (Mar. 13, 2014), http://www.thedailybeast.com/articles/2014/03/13/how-we-built-the-ghettos.html (last visited Oct. 7, 2020).

[36] *See*, *e.g.*, Nathalie Baptiste, *Them That's Got Shall Get,* AMERICAN PROSPECT (Oct. 12, 2014) *available at* http://prospect.org/article/staggering-loss-black-wealth-due-subprime-scandal-continues-unabated (last visited Oct. 7, 2020).

[37] *See* Sarah Burd-Sharps & Rebecca Rasch, Social Science Research Council, Report, *Impact of the U.S. Housing Crisis on the Racial Wealth Gap Across Generations* 2 (Jun. 2015), *commissioned by* ACLU, *available at* https://www.aclu.org/sites/default/files/field_document/discrimlend_final.pdf (last visited Oct. 7, 2020).

[38] *See* Andre M. Perry, *Black Americans were forced into 'social distancing' long before the coronavirus*, BROOKINGS (Mar. 20, 2020), *available at* https://www.brookings.edu/blog/the-avenue/2020/03/20/black-americans-were-forced-into-social-distancing-long-before-the-coronavirus/ (last visited Oct. 7, 2020) ("If we don't address the discrimination that is baked into current policy, efforts to address [COVID-19] will be undermined by the past practices that led to such inequality."); Gillian B. White, *The Recession's Racial Slant*, ATLANTIC (Jun. 24, 2015), *available at*

And modern redlining is a broader problem today than in the 1930s. For one, orthodox housing redlining occurs online - landlords have leveraged social media content moderation tools to keep apartment advertisements from showing online, but only in targeted zip codes.[39] MMTC spoke out against 'digital' redlining in 2011 when a black journalist discovered that Gmail advertisements changed based on the user's race[40] - if a White user sent an email about being arrested, Gmail displayed advertisements for fraud attorneys, but if a Black user sent the same note, Gmail displayed advertisements for DUI attorneys.[41] .

Provision of broadband access according to a locale's wealth distribution is another example of digital redlining. Due to America's racialized wealth distribution, it is less likely that racial minorities living under such schemes will have internet access.[42]

---

https://www.theatlantic.com/business/archive/2015/06/black-recession-housing-race/396725/ (last visited Oct. 7, 2020).

[39] *See generally* Summary of Settlements Between Civil Rights Advocates and Facebook (Mar. 19, 2019), *available at* https://www.aclu.org/other/summary-settlements-between-civil-rights-advocates-and-facebook (last visited Oct. 7, 2020) (summarizing settlement terms between Facebook and the ACLU regarding multiple instances of discriminatory advertising in housing, employment, and credit areas), *settling National Fair Housing Alliance v. Facebook*, No. 1:18-cv-02689 (S.D.N.Y. 2018).

[40] *See* Reniqua Allen, *Google Ads – Racial Profiling*, POLITIC365 (Oct. 2, 2011, *available at* http://politic365.com/2011/10/02/google-ads-%E2%80%94-racialprofiling/ (last visited May 21, 2016); *Racial Profiling in Your Own E-Mails?*, HISPANICALLY SPEAKING NEWS (Sept. 24, 2011), *available at* http://www.hispanicallyspeakingnews.com/notitas-denoticias/archives/2011/09/24/ (last visited May 21, 2016). *See also* Letter from Kim Keenan, President & CEO, MMTC, *to* Hon. Lawrence E. Strickling, Assistant Secretary for Communications and Information and Administrator, National Telecommunications and Information Administration, Regarding Racial Discrimination in Internet Commerce (Jan. 13, 2017), *available at* http://broadbandandsocialjustice.org/wp-content/uploads/2017/01/MMTC-App-Redlining-Ltr_011317.pdf (last visited Oct. 7, 2020) (hereinafter, "*App Redlining*") (referencing *application redlining*, which manifests when specific communities are structurally barred from using or being treated fairly on software applications).

[41] "Google denied these claims, but did admit that it read emails and used search history to determine which ads to display for each user." *App Redlining*, *supra* note 41, at 4.

[42] *Compare* Rakesh Kochhar & Richard Fry, Pew Research Center, *Wealth Inequality Has Widened Along Racial, Ethnic Lines Since End of Great Recession* (Dec. 12, 2014), *available at* http://www.pewresearch.org/fact-tank/2014/12/12/racial-wealth-gaps-great-recession (last visited Oct. 7, 2020) *with*, *e.g.*, John McQuaid, *Will Poor People Get Google Fiber?*, FORBES (Aug. 13, 2013), *available at* https://www.forbes.com/sites/johnmcquaid/2013/04/13/will-poor-people-get-

While such infrastructure planning is not ill-willed *per se*, it is part of a disturbing trend: in America, goods and services are provided by suppliers to consumers, and discrimination can occur when that provision follows a scheme justified by patterns derived from consumer data. The common denominator across all types of redlining is a person who loses equitable access to housing, apps, stores, and other basic tenants of civic society[43] – all because of who they are, and how that information fits into the suppliers' framework.

This is not to say that data-driven planning makes discrimination more likely.[44] Rather, it is to say that discrimination is a real and persistent societal trait, and it can be amplified when normal societal processes are boosted with data technology.[45] Modern data collection is clearly more advanced than data collection in the 1930s, but that difference may also create a false sense of social progress.[46] Scholars have opined that redlining was

---

google-fiber/#623f5f116d70 (last visited Oct. 7, 2020) (discussing racialized deployment of Google Fiber in Kansas City). *See also* Allana Akhtar, *Is Pokémon Go Racist? How the App May be Redlining Communities of Color*, USA TODAY (Aug. 9, 2016), *available at* http://www.usatoday.com/story/tech/news/2016/08/09/pokemon-go-racist-app-redlining-communities-color-racist-pokestops-gyms/87732734 (last visited Oct. 7, 2020) (noting "there were far more PokéStops and Gyms, locations where people pick up virtual goods or battle one another," in predominantly White neighborhoods than in minority areas).

[43] *See* Kwate et al., *Retail Redlining in New York City: Racialized Access to Day-to-Day Retail Resources*, *supra* note 21, at 632-33 (describing food, pharmacy, and banking deserts in low-income neighborhoods as symptoms of retail redlining).

[44] *See* Karthik Sivashanker et al., *A Data-Driven Approach to Addressing Racial Disparities in Health Care Outcomes*, HARVARD BUSINESS REVIEW (Jul. 21, 2020), *available at* https://hbr.org/2020/07/a-data-driven-approach-to-addressing-racial-disparities-in-health-care-outcomes (last visited Oct. 7, 2020) (describing how a public health center built a data infrastrtucture allowing them to, for example, "identify and escalate emerging risks to incident command leaders, such as our finding of higher mortality among non-English speaking Hispanic patients compared to English-speaking Hispanic patients."). *See also* Data for Black Lives, *supra* note 1 ("data activists can combine existing data on the racial makeup of counties & census tracts with [COVID] infection and death rates to better understand the impact on areas where Black people live.").

[45] *Accord* Introduction *in* Automated Intrusion, *supra* note 1 ("[O]utputs from algorithms and other automated tools can and do reinforce biases and lead to disparate results because the datasets they rely on reflect historical and existing discrimination.").

[46] *See also id.* ("Because [algorithmic] systems rely on large datasets and statistical analyses, their outputs are often perceived as neutral and not affected by biases in the same way as human decision-making.").

a form of social distancing imposed on minority communities well before 2020's COVID-19 pandemic.[47] Others have noted that the red lines originally denoting which neighborhoods were unworthy of home mortgages later outlined which neighborhoods COVID-19 hit the hardest.[48]

To prevent further cycles of data-driven discrimination (and its indirect consequences), big data actors should invest in defining, measuring, and detecting data trends that, if left unaddressed, could result in real-world discriminatory outcomes. That way, all communities can benefit from big data's planning power without fear of continued or worsened discrimination against minorities. That investment would also broadcast an awareness of social reality, and of the oft-ignored connection between technology development and civil rights as people experience them on a daily basis.

---

[47] *See* Perry, *Black Americans were forced into 'social distancing' long before the coronavirus*, *supra* note 38.

[48] *See*, *e.g.* Graziella Bertocchi et al., *COVID-19, Race, and Redlining*, IZA - INSTITUTE OF LABOR ECONOMICS (July 2020), *available at* http://ftp.iza.org/dp13467.pdf (last visited Oct. 7, 2020) ("[W]e uncover[ed] a persistence influence of the racial segregation induced by the discriminatory lending practices of the 1930s, by way of a diminished resilience of the black population to the shock represented by the COVID-19 outbreak.").

In 1965, a government formula was created to test whether voter suppression was so great as to warrant special voter protections in a given state.[49] Prior to then, anti-Black voter violence had been rampant in the United States since Black people gained the constitutional right to vote in 1850.[50] Pro-racist advocates, particularly in the South, used everything from arbitrary and unevenly applied literacy tests to full-out murder to discourage non-Whites from voting.[51] Thus when the Voting Rights Act was passed in 1965, it was as a triumph because it "provided for direct Federal action to enable blacks to register and vote."[52]

Through the VRA, the federal government created a two-step system to leverage available data to protect minority voting rights: first, a simple formula was applied to measure whether a state's voter suppression reached an unreasonable threshold[53]; second, new legal protections were applied to states where voter suppression – as measured under the formula – warranted federal intervention.[54] Such 'covered' jurisdictions included states that applied literacy tests, moral character assessments, and other arbitrary qualifications for the right to vote. If those states wanted to change voting requirements, they had to seek pre-clearance from the federal government first.[55]

Stated otherwise, the VRA identified, quantified, measured, and regulated discrimination by leveraging data concerning voting

---

[49] See THE VOTING RIGHTS ACT: SUMMARY AND TEXT 4, 12 (U.S. Comm'n on Civ. Rights ed., Sept. 1971) (last visited Oct. 7, 2020) ("VRA") (forbidding states from revoking a citizen's right to vote pursuant to performance or non-compliance with "any test or device.").

[50] *See* IBRAM X. KENDI, STAMPED FROM THE BEGINNING: THE DEFINITIVE HISTORY OF RACIST IDEAS IN AMERICA 234, 241-42, 244-46, 248-50 (2016).

[51] *Id.*

[52] VRA, *supra* note 49, at 4.

[53] *See* VRA § 4(a)-(b), *in id.* at 12 (applying federal protections in states where (1) the Attorney General determines that any unlawful 'test or device' was the basis for revoking a citizen's right to vote and (2) Census or election data show less than 50 percent voter turnout or registration in that state).

[54] *Id.*

[55] *See The Voting Rights Act: A Resource Page*, BRENNAN CENTER FOR JUSTICE (Aug. 4, 2015), *available at* https://www.brennancenter.org/analysis/voting-rights-act-resource-page (last visited Oct. 7, 2020) (stating the VRA required covered jurisdictions to "seek pre-approval of changes in voting rules that could affect minorities.").

and legislative patterns. This process was adjusted and supplemented numerous times by policymakers between 1965 and 2006,[56] but government officials still attempted to game the system. Local governments began manipulating district lines according to local racial demographics – a process popularly known as political gerrymandering.[57]

Indeed, racist actors, like anti-racist actors, were able to leverage population data, but for the nefarious purpose of diluting Black peoples' voting power despite the VRA's protection. Gerrymandering amplified when Census data became available - "detailed demographic information at the precinct level" enabled block-by-block analysis of citizens' race, as did subsequent software tools.[58] Decades of legal conflict ensued over political gerrymandering after the VRA, but not over the use of citizen data *per se* – rather, the national debate on gerrymandering focused on more Constitutionally-cognizable issues like political engineering and unlawful discrimination.[59] It follows that the act of data-driven redistricting persisted past the new millennium, continually impacting minorities' voting rights.[60]

But the entire conversation was forcibly shelved in 2013, when the Supreme Court voted 5-4 that the VRA's coverage formula was unconstitutional.[61] According to the Court's conservative

---

[56] *Id.*

[57] *See generally* Robert N. Clinton, *Further Explorations of the Political Thicket: The Gerrymander and the Constitution*, 59 IOWA L. REV. 1-48 (1973) (reviewing attempts to gerrymander in the years immediately following enactment of the VRA).

[58] *See* Kristen Silverberg, *The Illegitimacy of the Incumbent Gerrymander*, 74 TEX. L. REV. 913, 924 (1996).

[59] *See generally* Daniel Hayes Lowenstein, *You Don't Have to be a Liberal to Hate the Racial Gerrymandering Cases,* 50 Stan. L. Rev. 779 (1998) (analyzing legal controversies arising in the aftermath of redistricting efforts over the last few decades); Heather K. Gerken, *Understanding the Right to an Undiluted Vote*, 114 HARV. L. REV. 1663 (2001); Michael Parsons, *Clearing the Political Thicket: Political Gerrymandering for Partisan Advantage is Unconstitutional*, 24 WM. & MARY BILL RTS. J. 1107 (2016).

[60] *See generally* Bruce E. Cain & Emily R. Zhang, *Blurred Lines: Conjoined Polarization and Voting Rights*, 77 OHIO ST. L. REV. 867 (2016) (hereinafter, "*Blurred Lines*").

[61] About Section 5 of the Voting Rights Act, U.S. Dep't of Justice, *available at* https://www.justice.gov/crt/about-section-5-voting-rights-act (last visited Sept. 20, 2020) (citing *Shelby County v. Holder*, 570 U.S. 529, 553 (2013)) (The Court "held that it is unconstitutional to use the coverage formula in Section 4(b) of the Voting Rights Act to determine which jurisdictions are subject to the preclearance requirement of Section 5 of the Voting Rights Act.").

majority, the VRA had been successful in bolstering minority voting rights, and that continued reliance on the statute's preclearance approach, which was built on "decades-old data relevant to decades-old problems," no longer made sense.[62] Justice Ruth Bader Ginsburg famously compared the majority's reasoning to throwing away your umbrella because you are not getting wet.[63]

By rejecting a data-backed regulation, *Shelby County v. Holder* also blurred the contours of the voting rights issue – politicization followed the Court's rejection that the underlying discriminatory trends were no longer Constitutionally cognizable; it has become harder to "disentangl[e] partisan aims from racial motives" in the voting rights context without being able to lean on the legal importance of the data.[64] Ironically, *Shelby County's* most cognizable downstream effect might be the widespread unconstitutional voter suppression during COVID-19. [65] Regardless,

---

[62] *Shelby County*, 570 U.S. 553 (Roberts, J.).

[63] *Shelby County*, 570 U.S. 694 (Ginsburg, J., dissenting). *See also Blurred Lines*, *supra* note 60, at 871 (describing the coverage formula as "a crucial tool for scrutinizing and precluding the implementation of possible second-generation barriers to voting, such as new voter ID requirements and other laws that can make voting more difficult.").

[64] *See Blurred Lines* at 872 (describing how *Shelby County* made it more difficult in the voting rights context to distinguish partisan goals, racialized goals, and legitimate goals based on actual societal conditions).

[65] *See* Chad W. Dunn et al., *Legal Theories to Compel Vote-by-Mail in Federal Court*, Cal. L. Rev. (May 2020), *available at* http://www.californialawreview.org/theories-to-compel-vote-by-mail/ (last visited Oct. 7, 2020) (arguing that the VRA, *Bush v. Gore*, due process, and the 1st, 14th, 15th, and 26th Amendments are all bases to argue that vote-by-mail restrictions during the COVID-19 pandemic were unlawful). *See also Texas Democratic Party v. Abbott*, 2020 WL 2541971 (W.D. Tex., May 19, 2020) (holding that the existence and enforcement of a state law restricting mail-in voting was unconstitutional and likely an unlawful civil rights conspiracy), *stayed by* 2020 WL 2616080 (5th Cir., May 20, 2020) *and* 961 F.3d 389 (5th Cir., Jun. 4, 2020), *motion to vacate denied by* 140 S.Ct. 2015 (Jun. 26, 2020). *See also* 140 S.Ct. 2015 (Sotomayor, J., respecting) (noting that *Texas Democratic Party*'s 26th Amendment question should not be decided "in the context of an emergency application to vacate" the 5th Circuit's stay, but advising the 5th Circuit to "consider the merits of the legal issues in this case well in advance of the November election."). *See also id.* at 886-89 (discussing the new dynamic of data-justified discrimination). *See also* Mienah Z. Sharif et al., *Voter Disenfranchisement and COVID-19 are Connected — Here's Why*, SALON (Apr. 13, 2020), *available at* https://www.salon.com/2020/04/13/voter-disenfranchisement-and-covid-19-are-connected-heres-why) (last visited Oct. 7, 2020) ("Many state and local leaders are taking advantage of [the COVID-19 pandemic] to disenfranchise racial and ethnic minorities at the ballot box."); Jordan Ellenberg, *How Computers Turned Gerrymandering into a Science*, N.Y. TIMES (Oct. 6, 2017), *available at*

with one less federal mechanism to *actionize* voter suppression data, it has become tragically clear that voting rights is one area where big data can and must be leveraged to ensure human equity.[66]

---

https://www.nytimes.com/2017/10/06/opinion/sunday/computers-gerrymandering-wisconsin.html (last visited Oct. 7, 2020).

[66] *See* Nancy Abudu, *Seven years after Shelby County vs. Holder, voter suppression permeates the South*, SOUTHERN POVERTY L. CTR. (Jun. 25, 2020), *available at* https://www.splcenter.org/news/2020/06/25/seven-years-after-shelby-county-vs-holder-voter-suppression-permeates-south (last visited Oct. 7, 2020) ("Across the Southern states, in fact, you see a litany of polling place closures and photo ID requirements, restrictions on early voting, and efforts to block voting by mail in the midst of the COVID-19 pandemic, with people being forced to choose between their health – and that of their loved ones – and their constitutional rights. We are seeing a disparate impact on people of color.")

Data-driven policing has delivered decidedly mixed results for communities of color. Crime rates have lowered nationally in the past few years, and at the same time, more detailed policing data has become available.[67] Lower crime rates tend to increase community home values and attract outside investment,[68] which may incite calls for even more smart policing. Some civic leaders may feel similarly, especially in municipalities where police data is mutually accessible by officers and civilians.[69] Consistent with the above case studies, though, it must be explored whether use of data for law enforcement purposes produces equitably beneficial outcomes.

For example, crime statistics have justified the development and enforcement of metropolitan stop-and-frisk policies.[70] Communities of color were so disproportionately profiled and targeted that national outrage ensued in the early 2000s.[71] In

---

[67] The reliance on such data to improve policing outcomes on paper also created numerous new problems, including corruption in the collection of this information. *See, e.g.,* German Lopez, *This Policing Innovation Helped Fight Crime. But it Also Led to More Corruption*, Vox (Jul. 27, 2017), *available at* https://www.vox.com/2015/7/26/9036859/compstat-police-corruption (last visited Oct. 7, 2020).

[68] *See, e.g.,* Robert J. Shapiro & Kevin A. Hassett, *The Economic Benefits of Reducing Violent Crime: A Case Study of 8 American Cities,* Center for America Progress (Jun. 2012), *available at* https://cdn.americanprogress.org/wp-content/uploads/issues/2012/06/pdf/violent_crime.pdf (last visited Oct. 7, 2020).

[69] Daniel Nagin, *Police, Violence, and Data: The Black Lives Matter Movement*, CARNEGIE MELLON U., *available at* https://www.heinz.cmu.edu/media/2017/january/police-violence-data (last visited Oct. 7, 2020) (arguing that reforming police policy requires quality data "on what the police are doing in regard to the use of force," and "how communities feel about the police.").

[70] Report, Jeffrey Fagan, *NYPD Stop and Frisk Program Based on Race not Crime*, CENTER FOR CONSTITUTIONAL RIGHTS (Oct. 26, 2020), *available at* https://ccrjustice.org/home/press-center/press-releases/report-nypd-stop-and-frisk-program-based-race-not-crime (last visited Oct. 7, 2020) ("Fagan Report") (reporting that over half of the total stop and frisks occurring in New York City were motivated by "high crime area" status, "even in precincts with lower than average crime rates."). (Increased reliance on data by police departments also creates perverse incentives in the stop-and-frisk context vis-à-vis assuring that certain metrics are met each month lest crime be seen as increasing. *See, e.g.,* Ernie Naspretto, *How Stop-and-Frisk Went Off the Rails*, N.Y. DAILY NEWS (Jul. 30, 2013), *available at* http://www.nydailynews.com/opinion/stop-and-frisk-rails-article-1.1412159 (last visited Oct. 7, 2020).

[71] *See, e.g.,* Joseph Goldstein, *Judge Rejects New York's Stop-and-Frisk Policy*, N.Y. TIMES, (Aug. 12, 2013), *available at* http://www.nytimes.com/2013/08/13/nyregion/stop-and-frisk-practice-violated-

addition, in the vast majority of cases, these stops failed to yield any evidence of wrongdoing; that contributed to an unhealthy dynamic of distrust between those being stopped and the police.[72] The corresponding impact on crime is negative in light of the "demonstrated links between perceptions of fairness, evaluations of legitimacy, and police effectiveness."[73]

Alarmingly, even law enforcement leaders have scrutinized police use of data technologies – and not only data-driven police policymaking, but also the preceding data collection and analysis processes.[74] Police data entry systems are frequently poorly designed and badly used: first, when an officer enters information about an encounter into a police database, the quality of the data often depends on the quality of the officer's entry.[75] Second, after that information is entered, the quality of the data assessment depends on the quality of the system and the quality of the data being analyzed.[76] Given that such systems are used to justify and plan police practices, it is unsettling that – for many communities – neither the data being entered nor the analysis performed on it are

---

rights-judge-rules.html (last visited Oct. 7, 2020). *See also* Devan Cole et al., *'I was wrong': Bloomberg sorry for 'Stop and Frisk' in about-face apology ahead of potential presidential bid*, CNN (Nov. 17, 2019), *available at* https://www.cnn.com/2019/11/17/politics/michael-bloomberg-stop-and-frisk-apology/index.html (last visited Oct. 7, 2020).

[72] *See* Nancy G. La Vigne *et al.*, Urban Institute, *Stop and Frisk: Balancing Crime Control with Community Relations* (2014), *available at* http://www.urban.org/sites/default/files/publication/33661/413258-Stop-and-Frisk-Balancing-Crime-Control-with-Community-Relations.PDF (last visited Oct. 7, 2020).

[73] *Id.* at 19.

[74] *See, e.g.*, Transcript, James Comey, Director, FBI, *Hard Truths: Law Enforcement and Race* (Feb. 12, 2015), *available at* https://www.fbi.gov/news/speeches/hard-truths-law-enforcement-and-race (last visited Oct. 7, 2020) ("James Comey, *Hard Truths*") ("Not long after riots broke out in Ferguson late last summer, I asked my staff to tell me how many people shot by police were African-American in this country. I wanted to see trends. I wanted to see information. They couldn't give it to me, and it wasn't their fault. Demographic data regarding officer-involved shootings is not consistently reported to us through our Uniform Crime Reporting Program.").

[75] *Id.* ("Because reporting is voluntary, our data is incomplete and therefore, in the aggregate, unreliable.").

[76] *See* Danner, *supra* note 1 (discussing different quality metrics for different types of automated decision-making programs); Laney, *supra* note 11 (discussing different quality metrics for data).

reliable; it is difficult to imagine the resultant policy decisions would be any more sound.[77]

Nonetheless, data technology innovations still push policing into the future. One emerging trend is "predictive policing," where authorities analyze "various data sets" to "mov[e] law enforcement from reacting to crimes into the realm of predicting what and where something is likely to happen and deploying resources accordingly."[78] These efforts remain nascent in the United States, but there is evidence that discrimination is increasing in the few cities experimenting with predictive policing.[79] Many activists have different takes on the idea, though: some argue that data-driven law enforcement practices, like predictive policing, should be flatly banned;[80] others speaking within the Black Lives Matter movement argue to instead use data analysis to bolster community oversight of local police.[81]

Regardless of ones' stance on predictive policing or policing itself, it cannot be denied that data has enabled smarter, safer communities. Local leaders can and do use data to inform their

---

[77] *See especially* James Comey, *Hard Truths* ("I recently listened to a thoughtful big city police chief express his frustration with that lack of reliable data. He said *he didn't know whether the Ferguson police shot one person a week, one a year, or one a century*, and that in the absence of good data, 'all we get are ideological thunderbolts, when what we need are ideological agnostics who use information to try to solve problems.' He's right.") (emphasis added).

[78] *See* National Institute of Justice, *Predictive Policing, available at* https://www.nij.gov/topics/law-enforcement/strategies/predictive-policing/Pages/welcome.aspx (last visited Oct. 7, 2020).

[79] *See* Matt Stroud, *The Minority Report: Chicago's New Police Computer Predicts Crime, But is it Racist?*, VERGE (Feb. 19, 2014) , *available at* http://www.theverge.com/2014/2/19/5419854/the-minority-report-this-computer-predicts-crime-but-is-it-racist (Oct. 7, 2020) (discussing evidence of discrimination perpetuated by a Chicago predictive policing 'heat map').

[80] *See, e.g.*, Avi Asher-Schapiro, *California city bans predictive policing in U.S. first*, REUTERS (Jun. 24, 2020), *available at* https://www.reuters.com/article/us-usa-police-tech-trfn/california-city-bans-predictive-policing-in-u-s-first-idUSKBN23V2XC (last visited Oct. 7, 2020) (quoting Justin Cummings, Mayor, City of Santa Cruz, Cali.) ("Understanding how predictive policing and facial recognition can be disproportionately biased against people of color, we officially banned the use of these technologies in the city of Santa Cruz."). *But see* Nagin, *infra* note 81.

[81] *But see* Daniel Nagin, *Police, Violence, and Data: The Black Lives Matter Movement*, CARNEGIE MELLON U., *available at* https://www.heinz.cmu.edu/media/2017/january/police-violence-data (last visited Oct. 7, 2020) (arguing that reforming police policy requires quality data "on what the police are doing in regard to the use of force," and "how communities feel about the police."). *But see* Asher-Schapiro, *supra*.

advocacy – civic organizations partner with local law enforcement, for example, to provide more advanced and open civilian oversight programs.[82] Such an environment is simply not be possible without the data, technology, and a spirit of equity. Therein lies true innovation.

---

[82] *See*, *e.g.*, Megan Smith et al., *Using Technology and Data to Improve Community Policing: The Police Data Initiative*, WHITE HOUSE BLOG (Apr. 9, 2015) (describing several localities that "creat[ed] strong internal accountability and community feedback systems," and in doing so took advantage of "enormous opportunities which exist to analyze the rich data available on police/citizen encounters to improve police practices, community voice, and community trust."). *See also National Police Foundation Publishes Best Practices Guide for Police Open Data*, AP (Nov. 29, 2018), *available at* https://apnews.com/press-release/pr-businesswire/ed1198ecd0bc4c67bb698e955c644bde (last visited Oct. 7, 2020) (outlining progress from an Obama-era program wherein more than 140 law enforcement agencies "promot[e] the use of open data to encourage joint problem-solving, innovation, enhanced understanding, and accountability between law enforcement agencies and the communities they serve.").

Conclusion

As historically demonstrated, data can become a weapon used to legitimize discrimination. Left unaddressed, data-driven bias can harden discriminatory views and increase barriers to social and economic progress. The effects of such bias can harm a community for generations.

It is appealing to point to the objective nature of data, particularly where a positive outcome overshadows any negative elements. For example, open data, in the form of greater public access to stop, frisk, and arrest records, is proving to be a powerful tool in the fight to reform how the police interact with communities of color.[83] But much more progress remains to be made, and a reliance on new data sets and analytical techniques must be balanced with earnest attempts to forge trust in communities that have long suffered the profound consequences of discrimination in the policing and criminal justice context.

It is unfortunate that history might be repeating itself, but understanding the role of data in prior systemic discrimination will give us a better understanding of how to identify, measure, and prevent future discrimination – data driven or otherwise. As such, the lessons learned from previous eras should inform any policy response to big data. Vigilance and the political will to act will be critical to making real progress, as will ensuring that communities of color have a seat at the table for these discussions.

---

[83] *See, e.g.*, Abigail Margulis, *Advocacy Group Uses Data to Police the Police on Race*, CITIZEN-TIMES (Jan. 30, 2017), , *available at* http://www.citizen-times.com/story/news/local/2017/01/30/advocacy-group-uses-data-police-police-race/97050090/ (last visited Oct. 7, 2020).