Algorithms, Discrimination and the Law

MICHAEL SELMI*

Algorithmic decisionmaking has the potential to limit substantially the bias that influences so many human decisions. And yet the advent of algorithmic decisionmaking has been met by a torrent of criticism, particularly among legal scholars who highlight the potential ways in which algorithms may discriminate and express concern about the ability of the law to regulate discriminatory algorithms. This Essay confronts those criticisms and aims to show that algorithms are likely to be less discriminatory than human decisionmakers and that existing legal standards should prove adequate to evaluate algorithms. This is true in part because many algorithms are not the so-called black-box algorithms the critics focus on, and equally important, the disparate impact theory provides a critical and adequate vehicle for judicial review of algorithms, including those black-box algorithms.

TABLE OF CONTENTS

I. INTRODUCTION

Decisionmaking often involves predicting the future. An employer hiring an individual from among many applicants is predicting the future success of

* Foundation Professor of Law, Arizona State College of Law. Earlier versions of this Essay were presented at Arizona State College of Law, University of Pittsburgh Law School and the Annual Colloquium on Employment and Labor Law. I am grateful for the comments I received at those presentations as well as comments from Donald Braman, Stephanie Bornstein, Dimitry Karshedt, Pauline Kim, Mark Patterson, Erin Scharff, Josh Sellers, Charlie Sullivan, Steve Wilborn and terrific research assistance from Alex Davis and Ava Esler.

the individual, and implicitly, predicting that she would be better than (or at least as good as) other available applicants. When a financial institution makes a lending decision, it is making a prediction on whether the individual borrower will repay the loan. When lawyers accept cases, they are predicting how that case will likely be resolved. And in setting pretrial terms, a judge is predicting that the particular amount of bail or certain conditions will encourage an individual to show up for trial while deterring criminal activity until the trial.

Needless to say, our predictions often turn out to be wrong and our decisionmaking flawed, for a variety of well-documented and predictable reasons, ranging from emphasizing the familiar and the most recent data points regardless of how typical they might be, to making fundamental statistical mistakes.[1] Psychologists and behavioral economists have documented the systematic flaws in our thinking that can lead us to make poor decisions, often based on our limited experiences or our proclivities to see the world in limited ways.[2]

Nowhere is this more apparent than when it comes to discrimination. As almost goes without saying, decades of research have documented discriminatory results based on race, gender, national origin and other classes in employment, housing and the criminal justice system, the three areas I will discuss in this Essay.[3] To provide some common indicators: African-American men have unemployment rates twice that of white men regardless of prevailing economic conditions and suffer from substantial wage disparities.[4] Virtually

---

[1] An interesting recent article documented how experienced asylum judges, umpires, and mortgage lenders all fell prey to what is known as the "gambler's fallacy," a process by which individuals mistake the importance of short patterns. In these cases, the authors found that after granting asylum for several individuals in a row, or calling several strikes in a row, the individuals would in turn call a ball or deny asylum. *See* Daniel L. Chen, Tobias J. Moskowitz & Kelly Shue, *Decision Making Under the Gambler's Fallacy: Evidence from Asylum Judges, Loan Officers, and Baseball Umpires*, 131 Q.J. ECON. 1181, 1181–83 (2016). This is just one illustration of the many ways in which systematic errors occur.

[2] The literature is vast, for a sampling see generally JENNIFER L. EBERHARDT, BIASED: UNCOVERING THE HIDDEN PREJUDICE THAT SHAPES WHAT WE SEE, THINK, AND DO (2019); RICHARD H. THALER, MISBEHAVING: THE MAKING OF BEHAVIORAL ECONOMICS (2015); DANIEL KAHNEMAN, THINKING, FAST AND SLOW (2011), and DAN ARIELY, PREDICTABLY IRRATIONAL: THE HIDDEN FORCES THAT SHAPE OUR DECISIONS (2008).

[3] *See generally* Lincoln Quillian, Devah Pager, Ole Hexel & Arnfinn H. Midtbøen, *Meta-Analysis of Field Experiments Shows No Change in Racial Discrimination Over Time*, 114 PROC. NAT'L ACAD. SCIS. 10870 (2017) (employment discrimination); RICHARD ROTHSTEIN, THE COLOR OF LAW: A FORGOTTEN HISTORY OF HOW OUR GOVERNMENT SEGREGATED AMERICA (2017) (housing discrimination); Emma Pierson et al., *A Large-Scale Analysis of Racial Disparities in Police Stops Across the United States*, 4 NATURE HUM. BEHAV. 736 (2020) (discrimination in criminal justice).

[4] Even during the pandemic when so many jobs disappeared, African-Americans felt the brunt much more than whites. *See* Jonnelle Marte, *Gap in U.S. Black and White Unemployment Rates Is Widest in Five Years*, REUTERS (July 2, 2020), https://www.reuters.com/article/us-usa-economy-unemployment-race/gap-in-us-black-and-white-unemployment-rates-is-widest-in-five-years-idUSKBN2431X7 [https://perma.cc/Y9QP-GTFN].

every study on mortgage lending—and maybe it is all of them—have documented persistent discrimination in lending markets, whether it is from redlining certain neighborhoods, denying African-American borrowers loans that whites could obtain, or charging higher interest rates to minority borrowers.[5] Our criminal justice system is also rife with racial discrimination. From the first encounter with a police officer, to arrest, to bail setting, conviction and sentencing, African-Americans are treated more harshly than whites.[6]

As an illustration of how flawed our decisionmaking process can be, consider the ubiquitous employment interview. Research has shown that many interviewers make remarkably quick judgments, including assessing social class based on a brief interaction and placing undue emphasis on hobbies or other cultural markers found on a resume.[7] The unstructured interview—if you were

---

One of the most famous studies documenting racial discrimination in employment involved sending out resumes with different names, and those that were intended to be identifiably Black names (not an uncontroversial proposition) received strikingly fewer callbacks than those with white-sounding names. *See* Marianne Bertrand & Sendhil Mullainathan, *Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination*, 94 AM. ECON. REV. 991, 996–98 (2004). A recent evaluation of many field studies concluded that racial discrimination in employment had not decreased much over time. *See* Quillian, Pager, Hexel & Midtbøen, *supra* note 3, at 10870.

[5] The most prominent article, and one that sparked much of the subsequent research, originated at the Boston Federal Reserve Bank documenting discrimination in Boston based on government data. *See* Alicia H. Munnell, Geoffrey M. B. Tootell, Lynn E. Browne & James McEneaney, *Mortgage Lending in Boston: Interpreting HMDA Data*, 86 AM. ECON. REV. 25, 25–27 (1996). A recent study by Northwestern University scholars found that while overt discriminatory tactics such as racial steering had declined over the years, mortgage discrimination had not. *See* Lincoln Quillian, John J. Lee & Brandon Honoré, *Racial Discrimination in the U.S. Housing and Mortgage Lending Markets: A Quantitative Review of Trends, 1976–2016*, 12 RACE & SOC. PROBS. 13, 24–25 (2020). For an influential history of the role the federal government played in perpetuating segregation in housing see generally ROTHSTEIN, *supra* note 3.

[6] For a recent sampling of studies see Pierson et al., *supra* note 3, at 737–39; David Arnold, Will Dobbie & Crystal S. Yang, *Racial Bias in Bail Decisions*, 133 Q.J. ECON. 1885, 1889 (2018) ("We find three sets of facts suggesting that our results are driven by bail judges relying on inaccurate stereotypes that exaggerate the relative danger of releasing black defendants versus white defendants at the margin."); and U.S. SENT'G COMM'N, DEMOGRAPHIC DIFFERENCES IN SENTENCING: AN UPDATE TO THE 2012 BOOKER REPORT 2 (Nov. 2017) (documenting sentence disparities for identical crimes with identical backgrounds). For a recent gripping overview, see generally PAUL BUTLER, CHOKEHOLD: POLICING BLACK MEN (2017).

[7] *See* Michael W. Kraus, Brittany Torrez, Jun Won Park & Fariba Ghayebi, *Evidence for the Reproduction of Social Class in Brief Speech*, 116 PROC. NAT'L ACAD. SCIS. 22998, 22999 (2019) (finding that interviewers often make judgments about social class of those they interview after even just a few words of speech); Lauren A. Rivera & András Tilcsik, *Class Advantage, Commitment Penalty: The Gendered Effect of Social Class Signals in an Elite Labor Market*, 81 AM. SOCIO. REV. 1097, 1122 (2016) (finding that elite markers on a resume benefit men but not women); Lauren A. Rivera, *Hiring as Cultural Matching: The Case of Elite Professional Service Firms*, 77 AM. SOCIO. REV. 999, 1009 (2012) (finding that

an appliance, what kind of an appliance would you be kind of question—has been demonstrated repeatedly to be without any predictive validity and can actually impede accurate assessments.[8] Evidence has also long been clear that interviewers often seek to hire people like themselves.[9] Beyond the interview, references have been demonstrated to often be biased, particularly when it comes to gender, and the use of job referrals often excludes members of minority groups.[10] As one leading management scholar has recently concluded, when it comes to hiring, most companies are doing it all wrong.[11]

Given all that we know about human decisionmaking, particularly its biased nature, one might expect an eager embrace of alternatives that are designed to take the biased human out of the process, at least to the extent possible. Algorithmic decisionmaking is largely an attempt to do just that.[12] Algorithms have infiltrated many parts of our life, especially in consumer decisions with various nudges by Spotify, Amazon or Netflix designed to prompt us to buy

---

professional firms frequently treat hobbies listed on resumes as relevant and cultural markers).

[8] *See* Jason Dana, Robyn Dawes & Nathaniel Peterson, *Belief in the Unstructured Interview: The Persistence of an Illusion*, 8 JUDGMENT & DECISION MAKING 512, 519 (2013) ("In addition to the vast evidence suggesting that unstructured interviews do not provide incremental validity, we provide direct evidence that they can harm accuracy."); Scott Highhouse, *Stubborn Reliance on Intuition and Subjectivity in Employee Selection*, 1 INDUST. & ORG. PSYCH. 333, 333–34 (2008) (documenting inaccuracy of unstructured interviews despite persistent belief in their efficacy).

[9] This has been a widely documented phenomenon for decades. *See, e.g.*, Adam Grant, *What's Wrong with Job Interviews*, PSYCH. TODAY (June 11, 2013), https://www.psychologytoday.com/us/blog/give-and-take/201306/whats-wrong-job-interviews [https://perma.cc/TRW5-ACRW] ("Extensive research shows that interviewers try to hire themselves . . . ."); Greg J. Sears & Patricia M. Rowe, *A Personality-Based Similar-To-Me Effect in the Employment Interview: Conscientiousness, Affect- Versus Competence-Mediated Interpretations and the Role of Job Relevance*, 35 CANADIAN J. BEHAV. SCI. 13, 21–22 (2003); Thomas M. Rand & Kenneth N. Wexley, *Demonstration of the Effect, "Similar to Me," in Simulated Employment Interviews*, 36 PSYCH. REPS. 535, 541 (1975).

[10] Studies have demonstrated that recommendation letters tend to have stronger, more positive words for men than women. *See* Toni Schmader, Jessica Whitehead & Vicki H. Wysocki, *A Linguistic Comparison of Letters of Recommendation for Male and Female Chemistry and Biochemistry Job Applicants*, 57 SEX ROLES 509, 513 (2007). Recommendation letters for women seeking academic jobs tend to be more ambivalent regardless of the gender of the letter writer. *See* Juan M. Madera, Michelle R. Hebl, Heather Dial, Randi Martin & Virgina Valian, *Raising Doubt in Letters of Recommendation for Academia: Gender Differences and Their Impact*, 34 J. BUS. & PSYCH. 287, 287, 298 (2019). It has also long been documented that job referrals from existing employees tend to exclude individuals of different races. *See, e.g.*, David S. Pedulla & Devah Pager, *Race and Networks in the Job Search Process*, 84 AM. SOCIO. REV. 983, 983–85 (2019). Again, this is not a new observation. *See generally* Mark S. Granovetter, *The Strength of Weak Ties*, 78 AM. J. SOCIO. 1360 (1973).

[11] *See* Peter Cappelli, *Your Approach to Hiring Is All Wrong*, HARV. BUS. REV., May–June 2019, at 48, 50.

[12] *See infra* notes 22–23 and accompanying text.

more things and ideally to buy things that we may not have previously considered.[13] The use of algorithms have also grown substantially in mortgage lending, where a whole new industry known as FinTech has emerged to make decisions entirely online rather than with any in-person interaction.[14] In the criminal justice system, algorithms are frequently used for pretrial detention and sentencing determinations and have been used in some form for decades, depending on how one defines an algorithm.[15] And rather than relying on resumes or interviews, some employers are also embracing algorithms that might identify desirable employment characteristics based on piles of collected data, including from social media posts or from video interviews and games.[16]

Although the world has readily embraced algorithmic decisionmaking, legal scholars have been surprisingly critical of the development.[17] In what is now dozens of articles over the last few years, legal scholars have raised concerns about the discriminatory potential of algorithmic decisionmaking and relatedly the impotency of the law to address discrimination in the complex world of algorithms.[18] Both of these propositions seem contestable and I would suggest generally misguided.

---

[13] Mathias Jesse & Dietmar Jannach, *Digital Nudging with Recommender Systems: Survey and Future Directions*, COMPUTS. HUM. BEHAV. REPS. 1–2 (Jan. 13, 2021), https://www.sciencedirect.com/science/article/pii/S245195882030052X [https://perma.cc/UHL9-QSQH].

[14] There is a burgeoning literature on FinTech. *See, e.g.*, Sanjiv R. Das, *The Future of Fintech*, 48 FIN. MGMT. 981, 981–85 (2019); Andreas Fuster, Matthew Plosser, Philipp Schnabl & James Vickery, *The Role of Technology in Mortgage Lending*, 32 REV. FIN. STUD. 1854, 1854–55 (2019); William Magnuson, *Regulating Fintech*, 71 VAND. L. REV. 1167, 1174 (2018); Rory Van Loo, *Making Innovation More Competitive: The Case of Fintech*, 65 UCLA L. REV. 232, 238–40 (2018); Douglas W. Arner, Jànos Barberis & Ross P. Buckley, *The Evolution of FinTech: A New Post-Crisis Paradigm?*, 47 GEO. J. INT'L L. 1271, 1272–73 (2016).

[15] For a good overview of the use of risk assessments in sentencing see John Monahan & Jennifer L. Skeem, *Risk Assessment in Criminal Sentencing*, 12 ANN. REV. CLINICAL PSYCH. 489, 494–97 (2016).

[16] A number of employers have adopted a system created by HireVue that relies on artificial intelligence to analyze video interviews, including analyzing facial expressions and word usage. *See* Rachel Metz, *There's a New Obstacle to Landing a Job After College: Getting Approved by AI*, CNN (Jan. 15, 2020), https://www.cnn.com/2020/01/15/tech/ai-job-interview/index.html [https://perma.cc/VA7Z-433Z]. Many employers have also incorporated games into their hiring process. *See generally* Lydia Dishman, *Can Gamifying the Hiring Process Make It More Effective?*, FAST CO. (May 19, 2017), https://www.fastcompany.com/40422104/can-gamifying-the-hiring-process-make-it-more-effective [https://perma.cc/8KSF-YY2H].

[17] *See, e.g.*, Ifeoma Ajunwa, *The Paradox of Automation as Anti-Bias Intervention*, 41 CARDOZO L. REV. 1671, 1673–75 (2019).

[18] *See, e.g.*, *id.* at 1673–75; Pauline T. Kim, *Data-Driven Discrimination at Work*, 58 WM. & MARY L. REV. 857, 860–61 (2017); Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 673–75 (2016); Megan T. Stevenson & Christopher Slobogin, *Algorithmic Risk Assessments and the Double-Edged Sword of Youth*, 96 WASH. U. L. REV. 681, 681 (2018); Aziz Z. Huq, *Racial Equity in Algorithmic Criminal*

This first concern regarding the discriminatory potential of algorithms has an obvious component as well as a curious oversight. Many scholars have spent considerable time documenting the potential for algorithms to lead to discriminatory results, typically because the data they are based on are biased in some way.[19] This proposition seems entirely unremarkable, unless one thinks of algorithms as a magical process free of taint. As has been documented excessively, algorithms, no matter how they are ultimately created, depend on the data they analyze, and if those data are discriminatory, either by design or in their results, the ultimate algorithmic product will likely (though not necessarily) be discriminatory. In old computer lingo, this is known as garbage in, garbage out, which has more recently been translated to bias in, bias out.[20]

But again, this should not have been a revelation, and the real question of interest is not whether algorithms can produce discriminatory results—they can—but whether those results are likely to be more discriminatory than our existing systems. This question has been surprisingly ignored in the literature, and although a number of scholars will note that algorithms might exacerbate discrimination, most of the claims that algorithms can exacerbate or amplify discrimination remain more theoretical than real.[21] In contrast, a recent study regarding mortgage lending found that FinTech lenders led to less

---

Justice, 68 DUKE L.J. 1043, 1076–78, 1088 (2019); Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 6–8 (2014); Charles A. Sullivan, *Employing AI*, 63 VILL. L. REV. 395, 395–97 (2018); Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV. 633, 633 (2017); Anya E.R. Prince & Daniel Schwarcz, *Proxy Discrimination in the Age of Artificial Intelligence and Big Data*, 105 IOWA L. REV. 1257, 1259–61 (2020); Talia B. Gillis & Jann L. Spiess, *Big Data and Discrimination*, 86 U. CHI. L. REV. 459, 459–60 (2019); Ignacio N. Cofone, *Algorithmic Discrimination Is an Information Problem*, 70 HASTINGS L.J. 1389, 1412 (2019).

[19] *See* MEREDITH BROUSSARD, ARTIFICIAL UNINTELLIGENCE: HOW COMPUTERS MISUNDERSTAND THE WORLD 115 (2018) ("In an unequal world, if we make pricing algorithms based on what the world looks like, women and poor and minority customers inevitably get charged more."); Ajunwa, *supra* note 17, at 1699 ("Albeit that it is well documented that humans evince bias in employment decision-making, one cannot overlook that algorithmic systems of decision-making, too, might enable, facilitate, or amplify such biases." (footnote omitted)); Pauline T. Kim, *Manipulating Opportunity*, 106 VA. L. REV. 867, 869–70 (2020) ("Predictive algorithms . . . are likely to distribute information about future opportunities in ways that reflect existing inequalities and may reinforce historical patterns of disadvantage.").

[20] That phrase has found its way into two articles: Sandra G. Mayson, *Bias In, Bias Out*, 128 YALE L.J. 2218, 2224 (2019), and Ashesh Rambachan & Jonathan Roth, *Bias in, Bias Out? Evaluating the Folk Wisdom* 1 (Feb. 11, 2020) (unpublished manuscript), https://arxiv.org/abs/1909.08518 [https://perma.cc/G6R2-RJTD]. Engaging titles are the coin of the realm, particularly among books that are critical of algorithms. *See, e.g.*, CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY (2016); RUHA BENJAMIN, RACE AFTER TECHNOLOGY: ABOLITIONIST TOOLS FOR THE NEW JIM CODE (2019); VIRGINIA EUBANKS, AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR (2017).

[21] *See* Rambachan & Roth, *supra* note 20, at 1.

discriminatory results than traditional lenders.[22] Similarly, studies have indicated that judges fare worse than algorithms in bail or sentencing decisions, and a group of scholars have recently created an algorithm that significantly decreased discrimination in pretrial detention.[23] As a result, between an algorithm and a human, the smart money is likely on the algorithm.

Most of the algorithm critics readily acknowledge the discrimination that is rife in human decisionmaking but typically do so only in passing and their critiques carry a curious nostalgia for a simpler day, when an employer was willing to give that eager kid a chance, or when the local banker closed a deal with a firm handshake and a smile, or for that benevolent judge who let a teenager off with a stern lecture that led to a second chance. Some commentators are even explicit in their preference for discretion over the application of data analytics.[24] But it is all too easy to forget that those handshakes were typically among white men and that discretion is frequently a vehicle for discrimination, so much so that as early as the 1970s, courts began to take judicial notice of the connection between discrimination and discretion.[25] It should certainly be grounds for pause when scholars critiquing the discriminatory potential of algorithms retreat to a primary vehicle for perpetuating decades of discrimination, namely discretion.[26]

---

[22] Robert Bartlett, Adair Morse, Richard Stanton & Nancy Wallace, *Consumer-Lending Discrimination in the FinTech Era*, J. FIN. ECON. (forthcoming 2021) (manuscript at 26) (on file with the *Ohio State Law Journal*).

[23] *See* Jon Kleinberg, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig & Sendhil Mullainathan, *Human Decisions and Machine Predictions*, 133 Q.J. ECON. 237, 240–42 (2018) (detailing algorithmic risk assessment tool that decreased bias); Megan T. Stevenson & Jennifer L. Doleac, *Algorithmic Risk Assessment in the Hands of Humans* 1–2 (Hum. Cap. & Econ. Opportunity Glob. Working Grp., Working Paper No. 2020-055, 2020) ("[W]e find that racial disparities increased in the subset of courts where risk assessment appears most influential. This is partly . . . because judges were more likely to sentence leniently for white defendants with high risk scores than for black defendants with the same score.").

[24] *See* EUBANKS, *supra* note 20, at 81 ("Automated decision-making can change government for the better, and tracking program data may, in fact, help identify patterns of biased decision-making. But justice sometimes requires an ability to bend the rules.").

[25] Discretion in the form of subjective evaluations have long been linked to discrimination in the employment context. *See, e.g.*, Robinson v. Polaroid, 732 F.2d 1010, 1015 (1st Cir. 1984) ("[S]ubjective evaluations . . . could easily mask covert or unconscious race discrimination . . . ." (citing Rowe v. Gen. Motors Corp., 457 F.2d 348, 359 (5th Cir. 1972))); James M. Olson, Robert J. Ellis & Mark P. Zanna, *Validating Objective Versus Subjective Judgments: Interest in Social Comparison and Consistency Information*, 9 PERSONALITY & SOC. PSYCH. BULL. 427, 433 (1983).

[26] A frequent concern mentioned by critics is that an algorithm is likely to freeze out certain individuals who might otherwise have a chance in a human-driven process. *See* Ajunwa, *supra* note 17, at 1693 ("Whereas once, an applicant could rely on their interpersonal skills to make a favorable first impression on the hiring manager, these days the hiring algorithm is the initial hurdle to clear to gain employment."). This issue is discussed further in Part II.

In addition to demonstrating the potential for algorithmic discrimination, many of the critics have argued that the law will not be up to the task of regulating algorithms that produce discriminatory results.[27] Much of the concern involves the disparate impact theory that is available for challenges to both employment and lending decisions, though not criminal justice issues, and it has been argued that the inscrutable nature of algorithms will make any legal challenge ineffective.[28]

The concern regarding the limitations of existing legal standards are largely based on a failure to distinguish between two distinct types of algorithms: the first, what I refer to as Type 1 algorithms, are those that are understandable and are based on the vast quantities of available data, while the second kind of algorithm are variously referred to as inscrutable, opaque, or black-boxes because they are often difficult for humans to decipher (what I refer to as Type 2 algorithms). The first kind of algorithm presents no significant new problem for existing legal standards because courts have been adjudicating cases that involve complicated statistical procedures for decades.[29] When it comes to Type 2 algorithms, and contrary to the current consensus among the critics, I will demonstrate that it is more likely that defendants will be unable to defend algorithmic practices than that those practices will automatically be upheld. The disparate impact theory, which requires defendants to justify their practices under a business necessity test upon a showing of adverse impact, seems particularly well suited to adjudicate whether the algorithms survive legal scrutiny.[30] Moreover, the disparate impact framework allows plaintiffs to propose alternative practices, and here too, many algorithms should readily lend themselves to alternatives that would reduce disparate impact, and those that do not will likely not satisfy the prevailing business necessity tests. In other words, the concerns about the law's impotency seem overstated.

---

[27] *See, e.g.*, Kim, *supra* note 18, at 866 ("[A] mechanical application of existing disparate impact doctrine will fail to meet the particular risks that workforce analytics pose."); Gillis & Spiess, *supra* note 18, at 460 ("[W]e argue that legal doctrine is ill-prepared to face the challenges posed by algorithmic decisionmaking in a big data world."); Andrew D. Selbst, *Negligence and AI's Human Users*, 100 B.U. L. REV. 1315, 1372 (2020) ("Because disparate impact doctrine ties legitimate employment criteria to statistical predictions of future outcomes, properly executed machine learning models will often pass muster."); Sonia K. Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. REV. 54, 100 (2019) ("[O]ur existing statutory and constitutional schemes are poorly crafted to address issues of private, algorithmic discrimination.").

[28] The theory was established in the case of *Griggs v. Duke Power Co.*, 401 U.S. 424 (1971), and will be discussed further in Part III.A.

[29] The Supreme Court, in 1977, decided a trio of cases involving statistical analysis of discrimination. *See* Castaneda v. Partida, 430 U.S. 482, 485–88, 495–96 (1977) (grand jury); Int'l Brotherhood of Teamsters v. United States, 431 U.S. 324, 337–42 (1977) (employment discrimination involving long-haul drivers); Hazelwood v. United States, 433 U.S. 299, 310–13 (1977) (employment discrimination of teachers).

[30] *See infra* Part III.

This Essay will proceed as follows. Part Two will explore the nature of algorithms, explaining why only certain algorithms (Type 2) will pose any difficulty under existing legal standards and then will discuss the ways in which algorithms might discriminate, as well as how they also might be made to be less discriminatory than human actors. The third Part will apply existing legal standards to suggest that the disparate impact standard can adequately regulate even opaque algorithms and that some algorithms might even be challenged under the systemic discrimination model.

II. DEFINING ALGORITHMS AND DISCRIMINATION

A. *Defining Algorithms*

Although algorithms have only recently captured the attention of legal scholars, the concept has been a mainstay in computer science for decades, and courts have been adjudicating cases involving algorithms since at least the 1970s.[31] Today many people are familiar with algorithms as a result of the many consumer companies that rely on them to nudge people towards new purchases, the way, for example, Spotify seeks to introduce new music to listeners based on their past preferences as well as the related preferences and interests of other listeners.[32] There is, to be sure, a mysterious quality to algorithms, but in many ways, Spotify simply acts like a knowledgeable clerk at a record store nudging you to try new music typically based on the preferences of others with similar tastes.

This latter point—the comparison between the Spotify algorithm and a record store clerk—is important for understanding why it is that the law will have less difficulty parsing algorithms than is often asserted. Algorithms come in many flavors, and technically the term simply means a series of steps.[33] That series can be very basic (a recipe) or, as is increasingly common, can involve

---

[31] *See* Gottschalk v. Benson, 409 U.S. 63, 65 (1972) ("A procedure for solving a given type of mathematical problem is known as an 'algorithm.'"). Most of the early algorithm cases involved patent issues, particularly eligibility of software and math-type claims for patenting. *See, e.g.*, *In re* Application of Sarkar, 588 F.2d 1330, 1335 (C.C.P.A. 1978) (noting that "[i]f the steps of gathering and substituting values were alone sufficient, every mathematical equation formula, or algorithm having any practical use would be per se subject to patenting").

[32] *See* Jesse & Jannach, *supra* note 13, at 1–2.

[33] This is a common definition within the computer science literature. *See, e.g.*, BRIAN CHRISTIAN & TOM GRIFFITHS, ALGORITHMS TO LIVE BY: THE COMPUTER SCIENCE OF HUMAN DECISIONS 3–4 (2016) ("[A]n algorithm is just a finite sequence of steps used to solve a problem . . . . When you knit a sweater from a pattern, you're following an algorithm."). Similarly, one court has explained: "We note these discussions of the meaning of 'algorithm' to take the mystery out of the term and we point out once again that every step-by-step process, be it electronic or chemical or mechanical, involves an algorithm in the broad sense of the term." *In re* Iwahashi, 888 F.2d 1370, 1374 (Fed. Cir. 1989).

huge amounts of data that are analyzed in very little time.[34] But the fact that algorithms often have access to more data does not make them especially complicated or different from many traditional predictive devices, such as insurance policies. Insurance companies have always sought to use whatever relevant and legally permissible data were available to assess risk; the main difference now is that they have access to more data.[35]

Home mortgage lending offers another example. Over the last two decades mortgage lending has been transformed in two significant ways. First, the system has become more automated, and among some lenders, fully automated so that there is effectively no banker or broker shepherding the deal.[36] An entire industry known as FinTech has arisen that has captured a substantial portion of the lending market.[37] Even more traditional lenders now use automated underwriting, essentially algorithms that predict whether a borrower is likely to pay back their loan in a timely manner and to establish pricing based on the identified risk.[38] The mortgage industry is heavily regulated, and it is not particularly difficult to pierce the algorithms to know what factors are considered in making loan decisions.[39] As discussed in the next Part, litigation involving these systems has been common for many years, and although the cases themselves can be complicated, there is nothing particularly complicated about the way the law analyzes the lending decisions for discrimination.[40]

The second way in which the FinTech lenders differ is that they typically take in more data, and many of these lenders do so as a way of reaching a market untapped by traditional lenders who often require an extensive credit history that will disadvantage younger borrowers and those who may have untraditional financial histories, which until recently included many of those who were self-employed.[41] Because the data are now so readily available, lenders can incorporate factors such as bill paying that is not generally used by traditional lenders, and they may also look at information gleaned from social media.[42]

---

[34] *See* CHRISTIAN & GRIFFITHS, *supra* note 33, at 3–5.

[35] For an excellent history of risk and its application in the insurance industry see generally PETER L. BERNSTEIN, AGAINST THE GODS: THE REMARKABLE STORY OF RISK (1998).

[36] *See generally* J. CHRISTINA WANG, FED. RSRV. BANK OF BOS., TECHNOLOGY, THE NATURE OF INFORMATION, AND FINTECH MARKETPLACE LENDING (2018).

[37] A good overview can be found in *id.* at 36.

[38] *Cf. id.* at 14–15, 25–27.

[39] In their early stages, FinTech lenders listed their algorithms on their websites, and most of the factors they considered were virtually identical to those considered by traditional lenders. *See id.* at 10–11, 17.

[40] *See infra* Part III.A.

[41] *See* Sharon L. Lynch, *Self-Employed Turn to Non-Bank Lenders to Crack the Housing Market*, CNBC (Nov. 1, 2017), https://cnbc.com/2017/11/01/self-employed-turn-to-non-bank-lenders-to-crack-the-housing-market.html [https://perma.cc/H9DU-FDQ9].

[42] Many FinTech lenders target younger individuals without established credit histories and will often rely on data culled from social media. *See* Wang, *supra* note 36, at 12–13.

Facebook, for example, has developed a lending program that takes into account the credit quality of one's neighbors.[43]

The important thing to note is that incorporating more data into the decisionmaking or automating a process does not render these algorithms incomprehensible or beyond the reach of our existing legal structures. And most of the algorithms currently in use can be understood or analyzed in a way that makes them amenable to legal scrutiny. This is true of the very controversial algorithms currently in use to aid in bail setting and sentencing, algorithms that were intended in some jurisdictions to displace decades old checklists judges relied on.[44] We know that historically pretrial and sentencing decisions have been rife with discrimination, and there is some data to suggest that the algorithms have likewise produced discriminatory results.[45] But understanding what factors the algorithms take into account to determine whether an individual should be released before trial or what sentence she should receive is readily available. Both of the dominant companies in the market have, in fact, published the factors they consider, none of which is particularly innovative.[46] What the companies have typically declined to publish is what weights the factors receive, but in many cases this can be discerned through a process known as reverse engineering—something that law schools routinely do with the U.S. News & World Report rankings.[47]

---

[43] The program, which incorporates the credit ratings and other information from friends, is discussed in Robinson Meyer, *Could a Bank Deny Your Loan Based on Your Facebook Friends?*, ATLANTIC (Sept. 25, 2015), https://www.theatlantic.com/technology/archive/2015/09/facebooks-new-patent-and-digital-redlining/407287/ [https://perma.cc/9NM9-KYY4]. I use this example, which is not currently in use, as an indication of what companies are seeking to do, although obviously, basing lending decisions on one's neighbors could have discriminatory results given the prevalence of segregated housing.

[44] A report by ProPublica regarding the discriminatory results of a risk assessment tool known as COMPAS sparked much of the debate regarding algorithmic discrimination. The assessment tool was used for sentencing purposes and the analysis by ProPublica was based on data from Florida. For the initial report see Julia Angwin, Jeff Larson, Surya Mattu & Lauren Kirchner, *Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks.*, PROPUBLICA (May 23, 2016), https:/www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [https://perma.cc/RDH5-DJVY].

[45] The ProPublica article sparked a debate regarding what it meant for an algorithm to be discriminatory. The for-profit creators of the COMPAS risk assessment tool noted that it was equally accurate for whites and African-Americans in that it correctly predicted results about 68% of the time. *See id.* However, it was also determined that the particular errors went in the opposite direction—it overpredicted recidivism for African-Americans and underpredicted it for whites. The dispute over the COMPAS assessment tool has been written about extensively. For two concise discussions see Katyal, *supra* note 27, at 86–88, and Jessica M. Eaglin, *Constructing Recidivism Risk*, 67 EMORY L.J. 59, 95–98 (2017).

[46] *See* Stevenson & Slobogin, *supra* note 18, at 690–91 (discussing factors that are included in risk assessment tools).

[47] *See* Paul Caron, *Reverse Engineering the U.S. News Law School Rankings*, TAXPROF BLOG (Apr. 24, 2006), https://taxprof.typepad.com/taxprof_blog/2006/04/deconstructing_.html [https://perma.cc/JY3T-L7CK]. USNWR provides numerical factors for its calculations but

What is significant about these algorithms, and what runs contrary to much of the existing legal criticism, is they are readily amenable to legal scrutiny under existing standards. Most of those writing about algorithms, both legal scholars and others, often fail to distinguish among the various kinds of algorithms. Some even readily acknowledge that they are using the term to capture all forms of algorithms,[48] while others have a more idiosyncratic definition such as treating automated employment application systems as algorithms.[49] At the same time, the core focus of much of scholarship is on what are variously labelled black-box or inscrutable algorithms, Type 2 algorithms which currently comprise a relatively small portion of algorithms that are in use.[50] In the next Part, I will explain why I believe existing legal standards are adequate to evaluate claims of discrimination even by the black-box algorithms, but first it is important to explain how they differ from the more common Type 1 algorithm.

To understand how these black-box algorithms differ from something like a mortgage lending algorithm that analyzes huge amounts of data but in a decipherable way, we can use the likely familiar example of college admissions. A distinguishing feature of algorithms is that they are commonly trained on some past data set.[51] A mortgage lender for example will analyze past lending decisions and borrower behavior to determine what factors are relevant to

---

schools frequently reverse engineer the results of other schools to see where they might be able to make up ground in the rankings game. *See id.*

[48] *See, e.g.*, Cary Coglianese & David Lehr, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, 105 GEO. L.J. 1147, 1157 n.35 (2017) ("Although we explain some of these different terms in this Part, for the most part throughout this Article we use the terms 'machine learning,' 'algorithms' and 'artificial intelligence' for convenience to capture all possible variations in terms . . . ."). Others have likewise noted the increasingly liberal use of the term algorithm. *See* Robin K. Hill, *What an Algorithm Is*, 29 PHIL. & TECH. 35, 36 (2015) (noting that "we see evidence that any procedure or decision process . . . can be called an 'algorithm' in the press and in public discourse").

[49] Writing about the use of algorithms in employment hiring, Professor Ifeoma Ajunwa generally treats automatic processes as algorithms. *See* Ajunwa, *supra* note 17, at 1673 ("The automation of decision-making processes via machine learning algorithmic systems presents itself as a legal paradox."); *see also* Ifeoma Ajunwa & Daniel Greene, *Platforms at Work: Automated Hiring Platforms and Other New Intermediaries in the Organization of Work*, *in* WORK AND LABOR IN THE DIGITAL AGE 61, 62–63 (Steven P. Vallas & Anne Kovalainen eds., 2019).

[50] Even some of the most astute scholars assume that all algorithms are opaque. *See, e.g.*, Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan & Cass R. Sunstein, *Discrimination in the Age of Algorithms*, 10 J. LEGAL ANALYSIS 1, 2 (2019) ("[O]ne cannot determine what an algorithm will do by reading the underlying code."); *see also* Kim, *supra* note 18, at 881 ("[T]he resulting model is completely opaque, even to its creators."); Yavar Bathaee, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, 31 HARV. J.L. & TECH. 889, 892 (2018) ("Put simply, this means that it may not be possible to truly understand how a trained AI program is arriving at its decisions or predictions.").

[51] *See, e.g.*, Mayson, *supra* note 20, at 2224.

predicting repayment.[52] An ill-fated Amazon attempt to create an algorithm for future hires was based on data regarding its existing or past employees and based on that data the company then sought to identify employee characteristics that correlated with desirable workplace performance.[53] The project failed, but it was not a mystery what they were doing; what was a mystery, as discussed more below, is why Amazon did not anticipate some of the issues it encountered.

College admissions officers seek to accomplish a number of goals, including providing a meaningful and diverse experience for its students, but one of the goals is surely to select students who can perform at a high level. To fill an entering class, a college might rely on just a few factors: high school grade point average ("HSGPA") and standardized test scores perhaps. Even though there would be only two factors, one could still label this selection process as an algorithm since it would include some formula based around HSGPA and test scores. In using these two factors, the school would presumably base its predictions on how past students had performed, something that is both common and particularly important with respect to HSGPA since it can vary so much among secondary institutions.[54]

Rather than just looking at two variables, a college may choose to look at a large number of factors, which is what many colleges do and more say that they do.[55] These factors might include references, jobs, family income, or other factors the school deems relevant to academic performance, keeping in mind that for this example, the school is only trying to predict performance rather than assemble an interesting class (of course, the two need not be mutually exclusive). The school might also look at different factors, whether the student has a presence on TikTok, how many followers on Instagram, that sort of thing, and one of the critical elements of algorithms is they often will find unexpected patterns in the data.[56] Looking at its past students, a college may find that

---

[52] *See, e.g.*, Wang, *supra* note 36, at 10.

[53] The Amazon story has been oft told but it was originally reported by the news service Reuters. *See* Jeffrey Dastin, *Amazon Scraps Secret AI Recruiting Tool that Showed Bias Against Women*, REUTERS (Oct. 10, 2018), https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G [https://perma.cc/W794-PNEN].

[54] High school grades have typically been thought to be less valid predictors of college performance because of the wide variance among grades at schools, but a recent study suggested high school grades can be just as good a predictor as standardized tests. *See* Elaine M. Allensworth & Kallie Clark, *High School GPAs and ACT Scores as Predictors of College Completion: Examining Assumptions About Consistency Across High Schools*, 49 EDUC. RESEARCHER 198, 201, 209 (2020).

[55] For a rare and detailed look at the admissions process at an elite school see the opinions in the recent case unsuccessfully challenging Harvard's admission as discriminatory against Asian Americans, *Students for Fair Admissions, Inc. v. President and Fellows of Harvard College*, 397 F. Supp. 3d 126, 133–54 (D. Mass. 2019), affirmed in 980 F.3d 157 (1st Cir. 2020).

[56] *See* Huq, *supra* note 18, at 1066 (noting that machine learning will "generat[e] utterly unexpected outcomes"); Rob Kitchin, *Big Data, New Epistemologies and Paradigm Shifts*, BIG DATA & SOC'Y 2 (Apr. 1, 2014), https://journals.sagepub.com/doi/pdf/10.1177

whether a student had taken a computer science course in high school was strongly correlated with academic success in college regardless of their major, or they may find value in something more esoteric such as whether the person lacked a Facebook page. In a study of employees, for example, it was determined that there was a correlation between those new employees who downloaded their own browser when they started employment and workplace success.[57]

These kinds of algorithms—those that effectively combine big data with sophisticated statistical techniques—are now ubiquitous in our lives, and they often seek to find patterns in data that would take humans years to find. However, and this is important particularly for legal analysis, in most cases, if they had the time, humans could find the patterns. The mystery is not so much in the construction of the algorithm but rather the speed at which it is constructed. It is also the case that we may not understand why the algorithm identified the particular pattern or, on first glance, how important that factor is but with some work, in these circumstances, the underlying construct of the algorithm can generally be understood, and what will turn out to be important for legal analysis functions primarily like a regression.[58] Although I will return to this issue in the last Part, it is perhaps worth noting that courts, including the Supreme Court, have been evaluating regressions in the context of discrimination cases for more than forty years.[59]

There is, however, a kind of algorithm that does prove mysterious and that may be inscrutable by humans in some circumstances. This is what I am referring to as a Type 2 algorithm and what is often defined as an unsupervised machine learning algorithm, although it turns out that terms within computer

/2053951714528481 [https://perma.cc/6C4X-BRV4] ("Big Data analytics enables an entirely new epistemological approach for making sense of the world [and] rather than testing a theory by analyzing relevant data, new data analytics seek to gain insights 'born from the data.'").

[57] *See* Joe Pinsker, *People Who Use Firefox or Chrome Are Better Employees*, ATLANTIC (Mar. 16, 2015), https://www.theatlantic.com/business/archive/2015/03/people-who-use-firefox-or-chrome-are-better-employees/387781/ [https://perma.cc/H7FL-546N]. Contrary to the headline, the key finding was not which browser someone used for the assignment but rather that the employees downloaded their own browser rather than using the default.

[58] A number of authors have made the connection between big data algorithms and regressions. *See, e.g.*, Coglianese & Lehr, *supra* note 48, at 1156–58 (discussing similarities between regression equations and algorithms); Bathaee, *supra* note 50, at 900 (same).

[59] *See generally* Bazemore v. Friday, 478 U.S. 385 (1986) (per curiam) (analyzing regression equations used in pay discrimination case). The following year, the Supreme Court also considered a regression analysis in a death penalty case. *See* McCleskey v. Kemp, 481 U.S. 279, 327 (1987).

science vary considerably.[60] Regardless of the label, the way the algorithm is constructed can be explained.[61]

Returning to college admissions, the college will provide all of the data it has collected and then specify a desired outcome—for example, the college wants the most academically promising class of 2000 students selected from among an applicant pool of say 10,000. In addition to all the data on the applicant pool, the program will also have access to data on past students and perhaps data on students from other institutions as well. And then, and this is where the difference with other kinds of algorithms comes in, the college will effectively tell the algorithm to compile the class. In an unsupervised learning mode, the algorithm will identify the most promising 2,000 students, but it will not explain why those students were chosen, and it may be difficult for someone to understand what the decision process was and why these particular students were chosen compared to those who were not. In fact, if the algorithm could talk, and it was asked what the selection process was, it may be unable to articulate all of the details, though it could likely explain the process.

This is apparently what Amazon sought to create several years ago when it assembled a group of programmers to create an algorithm that would identify future successful employees.[62] Trying to identify successful employees is a notoriously difficult proposition and the sheer volume of applications companies receive make individualized attention all but impossible. As a result, Amazon wanted to create an algorithm that would identify successful employee characteristics, and it apparently was planning to use the algorithm to go out and locate individuals with those traits rather than parsing through piles of applications.[63] The algorithm was trained on data relating to past employees, both those who had been successful and those who had not, and the algorithm was then asked to choose among a group of "applicants" based on the correlations it had discovered in the data.[64] Things went seriously wrong for what should have been a fairly predictable reason: men were overwhelmingly selected and applicants were apparently penalized for gender specific markers on a resume such as "women's chess club."[65]

The Amazon story has received widespread attention, but it highlights an important aspect of algorithms—they are trained on data, they are not magic formulas, and if the data are skewed in some fashion, the results may be too. But

---

[60] For a discussion of the various terms that are used see ED FINN, WHAT ALGORITHMS WANT: IMAGINATION IN THE AGE OF COMPUTING 1–2, 28–35 (2017).

[61] For purposes of this Essay, I do not think it is necessary to provide the technical details of how these black-box algorithms are constructed. For those who are interested there are many accessible resources both within and outside of law. *See generally* BROUSSARD, *supra* note 19; CHRISTIAN & GRIFFITHS, *supra* note 33; SEAN GERRISH, HOW SMART MACHINES THINK (2018).

[62] *See* Dastin, *supra* note 53.

[63] *Id.*

[64] *Id.*

[65] *Id.*

as discussed in the next Part, even when the training data are biased, that does not necessarily mean that the algorithm will produce discriminatory results. Another key and often neglected aspect of the Amazon story is that the company was presumably able to identify what went wrong in the algorithm, and it quickly abandoned the project, never using it for actual hiring.

B. *Defining Discrimination*

The Amazon case reflects how algorithms can discriminate. It does not appear that anyone at Amazon set out to establish a hiring algorithm that favored men, but it was instead dependent on the data it had.[66] Although the Amazon example provides a cautionary tale, it should also provide concern to the critics—the reason Amazon's algorithm produced discriminatory results was because Amazon's hiring practices, run by humans, led to a predominantly male workforce.[67] The humans caused the machine to discriminate. And recent research has revealed that humans can also likely cause the machine not to discriminate, an issue I will return to momentarily.[68]

Algorithm critics emphasize that discrimination is likely to occur not by direct means but by proxies.[69] Currently, algorithms for employment, lending, or the criminal justice system omit direct markers of race and gender but there is a concern that an algorithm might discriminate based on things like addresses or arrest records that may correlate with race, gender, or some other prohibited category.[70] This, however, seems an unusual critique, and one that overlooks the long history of litigation challenging the use of proxies in various settings.

One of the reasons the critique seems unusual is that modern discrimination commonly works through what the algorithm literature now defines as "proxies" and has for many decades. For example, there were successful challenges to the use of arrest records in employment beginning in the 1970s when several courts struck down the use of arrest records because of their unjustified effect on

---

[66] *Id.* ("In effect, Amazon's system *taught itself* that male candidates were preferable." (emphasis added)).

[67] *See id.* ("Most [resumes] came from men, a reflection of male dominance across the tech industry.").

[68] *See* Rambachan & Roth, *supra* note 20, at 1.

[69] *See, e.g.*, Kim, *supra* note 18, at 898 ("Because other information contained in large datasets can serve as a proxy for race, disability, or other protected statuses, simply eliminating data on those characteristics cannot prevent models that are biased along these dimensions.").

[70] *See, e.g.*, *id.* at 877 ("Data models may also discriminate when neutral factors act as 'proxies' for sensitive characteristics like race or sex. Those neutral factors may be highly correlated with membership in a protected class, and also correlate with outcomes of interest."); Barocas & Selbst, *supra* note 18, at 691 (emphasizing that discrimination can arise based on "reliable proxies for class membership"); Kleinberg, Ludwig, Mullainathan & Sunstein, *supra* note 50, at 4 (noting that discrimination might arise based on "past arrest records [that] are used to predict the likelihood of future crime"). For a discussion of proxy discrimination in the insurance context, see generally Prince & Schwarcz, *supra* note 18.

African-Americans.[71] More recently, many jurisdictions have adopted what are known as ban-the-box laws, which typically restrict employers' use of arrest records in the application phase.[72] Address-based discrimination can have a similar effect. After all, redlining is a form of proxy discrimination when lenders effectively wall off entire neighborhoods, and residency requirements can be seen in the same light—they restrict hiring to residents of a particular municipality while excluding others and, because of our segregated neighborhoods, such policies can often be discriminatory.[73] Just as was true with arrest records, successful challenges to the residency requirements of majority white jurisdictions arose in the 1980s and continued for many years thereafter.[74] Similarly, a highly influential study by economists documented that individuals with identifiably white names on their resumes received substantially more call-

---

[71] Challenges to arrest records originated in the very early days of Title VII. *See, e.g.*, Gregory v. Litton Sys., Inc., 316 F. Supp. 401, 402 (C.D. Cal. 1970) ("There is no evidence to support a claim that persons who have . . . no criminal convictions but have been arrested . . . can be expected . . . to perform less efficiently or less honestly than other employees."), *aff'd as modified*, 472 F.2d 631 (9th Cir. 1972). Despite the early successful cases, the challenges continue today, and are also typically successful. *See, e.g.*, Conn. Fair Hous. Ctr. v. CoreLogic Rental Prop. Sols., 478 F. Supp. 3d 259, 300 (D. Conn. 2020) (holding that there was no business justification for considering arrest as part of rental screening process).

[72] Thirty-six states have now adopted some form of ban-the-box legislation. *See* BETH AVERY & HAN LU, NAT'L EMP. L. PROJECT, BAN THE BOX 2 (Oct. 2020), https://www.nelp.org/publication/ban-the-box-fair-chance-hiring-state-and-local-guide/ [https://perma.cc/MH5W-RUUM]. Ironically, but related to our flawed systems, several studies have indicated that such legislation can increase discrimination because employers are likely to assume criminal records for African-American candidates. *See, e.g.*, Jennifer L. Doleac & Benjamin Hansen, *The Unintended Consequences of "Ban the Box": Statistical Discrimination and Employment Outcomes when Criminal Histories Are Hidden*, 38 J. LAB. ECON. 321, 321 (2020) (concluding that ban-the-box legislation decreases the probability of employment by 5.1% (3.4 percentage points) for young low-skilled African-American men); Harry J. Holzer, Steven Raphael & Michael A. Stoll, *Perceived Criminality, Criminal Background Checks, and the Racial Hiring Practices of Employers*, 49 J.L. & ECON. 451, 451 (2006) (finding that employers who conducted criminal background checks were more likely to hire African-Americans).

[73] Prince & Schwarcz, *supra* note 18, at 1262 ("Indeed, the paradigmatic example of proxy discrimination by humans involves financial firms that refused to serve predominantly African American geographic regions, a phenomenon known as redlining."); *see also* NAACP v. N. Hudson Reg'l Fire & Rescue, 665 F.3d 464, 481 (3d Cir. 2011) (explaining that city's residency requirement "creates a disparate impact on African American firefighter applicants").

[74] Many of the cases were brought by the NAACP against jurisdictions in New Jersey. *See, e.g.*, *N. Hudson Reg'l Fire & Rescue*, 665 F.3d at 468; Newark Branch, NAACP v. Town of Harrison, 940 F.2d 792, 794 (3d Cir. 1991); Newark Branch, NAACP v. Township of West Orange, 786 F. Supp. 408, 411, 434 (D.N.J. 1992). The Civil Rights Division of the Justice Department also initiated a series of cases in the suburbs surrounding Detroit, all of which settled but one. *See* United States v. City of Warren, 138 F.3d 1083, 1088 (6th Cir. 1988).

back offers than those with identifiably African-American names, a finding that is a form of proxy discrimination.[75]

What this brief analysis demonstrates is that concern with what the algorithm literature labels proxy discrimination is neither new nor unique to algorithms. Human decisonmakers have relied on proxies to discriminate for likely as long as they have been discriminating. One significant advantage to an algorithmic process is that the data can be withheld so that an algorithm would not have access to race, gender, addresses, arrest records or other potential forms of discrimination.[76] This does not mean that an algorithm will never produce discriminatory results because of tainted variables, but it does mean that, again, algorithms are unlikely to be more discriminatory than humans, and there is good reason to believe they will be less.

On the negative side, an algorithm may identify correlations that prove to be discriminatory, and given the nature of algorithms it may be more difficult, and with Type 2 algorithms perhaps impossible, to uncover the variables that are functioning as a proxy for discrimination.[77] Given just how racially and gender stratified our society is, there are many potential factors that could reflect race and gender and which an algorithm might take into account.[78] Both of these issues will be discussed in the next Part, but for Type 1 algorithms where the underlying code can be revealed in some fashion, there is nothing distinctive between an algorithm and other issues that have been litigated under the law in housing and employment.

The 1986 case of *Bazemore v. Friday* provides an illustration.[79] The case involved pay discrimination between white and African-American agricultural extension agents, state employees who helped farmers with their crops.[80] In many ways the case was a standard pay discrimination case where regression analyses were used to identify what explained the observed pay differences between African-American and white agents.[81] Initially, all that was known was that African-Americans were paid less than their white counterparts but it was

---

[75] Bertrand & Mullainathan, *supra* note 4, at 991. More recently, a similar study found similar results. *See* Patrick M. Kline, Evan K. Rose & Christopher R. Walters, *Systemic Discrimination Among Large U.S. Employers* 3 (Nat'l Bureau Econ. Rsch., Working Paper No. 29053, 2021), https://eml.berkeley.edu//~crwalters/papers/randres.pdf [https://perma.cc/59WH-45U4].

[76] In some circumstances it may also be possible to keep defining information away from humans, as ban-the-box statutes attempt to do, and perhaps the best known example of such an attempt is orchestras moving to hold auditions behind partitions. *See generally* Claudia Goldin & Cecilia Rouse, *Orchestrating Impartiality: The Impact of "Blind" Auditions on Female Musicians*, 90 AM. ECON. REV. 715 (2000).

[77] *See* Bathaee, *supra* note 50, at 891.

[78] *See, e.g.*, Kim, *supra* note 18, at 898.

[79] Bazemore v. Friday, 478 U.S. 385 (1986) (per curiam).

[80] *Id.* at 390–91 (Brennan, J., concurring in part).

[81] *Id.* at 398.

not known why.[82] As it turns out, a key explanatory factor was what crop the agents were responsible for—in particular, those who had responsibility for tobacco were paid more, something that in a tobacco state like North Carolina was not all that surprising.[83] But it also turned out that virtually all of the agents who were assigned to the tobacco crop were white, which raised the question whether discrimination explained either the pay or assignment.[84] In the context of the litigation, if only whites were assigned to the most valuable North Carolina crop, then it would not be appropriate to include crop assignment as a factor in the regression analysis that was designed to analyze the pay differential because the crop assignment would be a proxy for discrimination, just as addresses, arrest records or other markers might be. This was not a new issue in 1986 and it is not a new issue now—any variable that is tainted by discrimination should be excluded (or corrected where possible) from the analysis.

A number of scholars have raised a related but slightly different concern, namely that the use of algorithms as selection devices will overlook outliers, those who do not fit the characteristics identified by the algorithm.[85] In her important work, Cathy O'Neil noted, for example, that an automated process might disadvantage those with disabilities, and Professor Pauline Kim has recently raised concerns that an algorithm might exclude many qualified individuals from the outset, never allowing them an opportunity to be reviewed.[86] This issue, however, again has little to do with algorithms and all to do with using data or statistical analyses, which emphasize averages rather than individual considerations, essentially a variation on the old rules versus standards debate. The same issue arises with any screening device whether it is in the form of an algorithm, the use of standardized tests, or even something like a minimum age requirement. Any and all of these devices will necessarily

---

[82] *Id.* at 403 ("[P]etitioners presented evidence consisting of individual comparisons between salaries of blacks and whites similarly situated. Witness testimony . . . also confirmed the continued existence of such disparities.").

[83] Michael Selmi, *Statistical Inequality and Intentional (Not Implicit) Discrimination*, 79 L. & CONTEMP. PROBS. 199, 214 & n.77 (2016).

[84] In the Supreme Court, the *Bazemore* case involved a number of complicated issues, including whether the state was obligated to cure discrimination that had occurred prior to when the statute became applicable to public employers in 1972. *See Bazemore*, 478 U.S. at 394. The Supreme Court also concluded that the District Court had erred when it rejected the plaintiff's regression analyses because they had not considered all possible variables that might explain the observed salary disparities. *Id.* at 400. On remand, additional analyses indicated that crop assignments were a significant explanatory factor, although the case settled before a retrial.

[85] *See, e.g.*, O'NEIL, *supra* note 20, at 105–12.

[86] Cathy O'Neil focuses on a college student named Kyle Behm who had bipolar disorder and who repeatedly failed the same psychometric screening examination for a low-level job, adding, "[u]nder the previous status quo, employers no doubt had biases. But those biases varied from company to company, which might have cracked open a door somewhere for people like Kyle Behm." *Id.* at 112; *see also* Kim, *supra* note 19, at 874 (emphasizing how bias can screen out potential applicants "before they even interact with a hiring firm").

exclude some qualified individuals, which can be problematic if the exclusion is based on a prohibited basis but is less concerning under the law when it does not. There is also an implicit assumption in these critiques that a non-algorithmic process would select the outliers even though we know that is unlikely to occur.

Many of the issues relating to the discriminatory potential of algorithms are not unique to algorithms but are problems that have been a staple of antidiscrimination law.[87] To date, there is little evidence that algorithms are likely to be more discriminatory than existing systems, although there is some basis to believe they can be less discriminatory.[88] The way the issue has unfolded, it seems that the explosion of interest in algorithms over the last decade did not initially prioritize an antidiscrimination ethic. While most people within the computer science profession were aware that algorithms might produce troubling and discriminatory results, they were primarily concerned with ensuring accuracy rather than equity.[89] In the last few years, that has changed and today there is essentially a whole subfield devoted to making algorithms accurate and fair, and ideally, more fair than our existing systems.[90]

Two recent articles provide insight into how algorithms can be designed to pursue equity goals.[91] In a recent paper, Professor Bo Cowgill has demonstrated how it is possible to exploit the noise in algorithms to effectively debias the underlying data.[92] As Professor Cowgill explains and demonstrates through a theoretical discussion, "[d]epending on the level of noise, an algorithm can either replicate historical human bias or completely correct it."[93] He later concedes that completely correcting the bias in the data is unlikely to be a

---

[87] *See* Arnold, Dobbie & Yang, *supra* note 6, at 1889 (explaining that "bail judges rely[] on inaccurate stereotypes that exaggerate the relative danger of releasing black defendants versus white defendants").

[88] *See* Rambachan & Roth, *supra* note 20, at 1.

[89] For a discussion regarding the evolution of research see generally MICHAEL KEARNS & AARON ROTH, THE ETHICAL ALGORITHM: THE SCIENCE OF SOCIALLY AWARE ALGORITHM DESIGN (2020).

[90] *See generally id.*; Bo Cowgill & Catherine Tucker, *Algorithmic Fairness and Economics* (Feb. 14, 2020) (unpublished manuscript), https://ssrn.com/abstract=3361280 (on file with the *Ohio State Law Journal*); Alekh Agarwal, Alina Beygelzimer, Miroslav Dudík, John Langford & Hanna Wallach, *A Reductions Approach to Fair Classification*, PROC. MACH. LEARNING RSCH. (2018), http://proceedings.mlr.press/v80/agarwal18a/agarwal18a.pdf [https://perma.cc/Q7T4-W5RJ]; Michael Veale & Reuben Binns, *Fairer Machine Learning in the Real World: Mitigating Discrimination Without Collecting Sensitive Data*, BIG DATA & SOC'Y (Nov. 20, 2017), https://journals.sagepub.com/doi/pdf/10.1177/2053951717743530 [https://perma.cc/6DGN-RA59].

[91] *See generally* Bo Cowgill, *Bias and Productivity in Humans and Machines* (Upjohn Inst., Working Paper No. 19-309, 2019), https://ssrn.com/abstract=3433737 [https://perma.cc/Y4KA-VLE3]; Rambachan & Roth, *supra* note 20.

[92] Cowgill, *supra* note 91, at 1.

[93] *Id.* at 2. By noise, Professor Cowgill means the "random extraneous factors in human decision-making" that by definition "are not predictive of the candidate's underlying quality." *Id.* at 6–8. The computational means for exploiting the noise is complicated, but the implication is important—biased data need not lead to biased algorithms. *Id.*

realistic goal, but the important point is that it is possible to exploit the algorithm to reduce bias, largely because the underlying data is likely highly inaccurate.[94] This is just one way researchers are learning to reduce bias with algorithms and given the interest in the topic, it is likely that other means of reducing bias will be developed.

Another recent paper offers additional insights that suggest algorithms may produce less discriminatory results than would arise from human decisionmaking.[95] As an example, two Harvard professors rely on a well-known phenomenon involving police stops and searches.[96] It has been widely documented that African-Americans are stopped by police far more frequently than white individuals but that contraband is found more frequently among those white individuals who are stopped.[97] As a result, an algorithm that was designed to produce the most effective police stops would encourage the stopping of fewer African-Americans, thus reducing the discriminatory effect.[98] Similarly, algorithms designed for an employment setting might also identify characteristics that are correlated with minority or female employees. In many predominantly white or male workplaces, the minority or female employees are likely to be some of the best employees because they will likely have to overcome institutional barriers that may not be present for white men, and these individuals may provide a basis for obtaining more minority and female workers, just as was true in the police searches whereby fewer African-Americans should be stopped if data guided police behavior.

These insights run counter to the assumption of the critics that biased data will lead to biased results. On the contrary, the reverse is possible. As the authors explain:

> A biased hiring manager applies a higher predicted-productivity threshold for African Americans than for whites . . . . Thus, the more biased is the hiring manager against African Americans, the higher will be the algorithm's predicted productivity for African Americans and the more African Americans will be hired by an algorithmic hiring rule.[99]

In reading this explanation, one might wonder how affirmative action would affect the analysis, but if one's first instinct is to assume that African-Americans hired into a predominantly white workforce are the product of affirmative action, then one can see how stereotypes affect human decisionmaking in a way that an algorithm can avoid.

---

[94] *See id.* at 2, 18. For a book length treatment on the concept of noise and how pervasive it is in decisionmaking see DANIEL KAHNEMAN, OLIVER SIBONY & CASS R. SUNSTEIN, NOISE: A FLAW IN HUMAN JUDGMENT (2021).

[95] *See generally* Rambachan & Roth, *supra* note 20.

[96] *Id.*

[97] Studies uniformly show both the more frequent stops among African-Americans and less frequent hits. I discuss the studies in Selmi, *supra* note 83, 207–12.

[98] *See* Rambachan & Roth, *supra* note 20, at 2.

[99] *Id.* at 7.

An additional and likely the easiest way to reduce the bias would be to exclude biased data from the algorithm. As noted previously, this is already commonly done—race and gender are typically excluded, addresses and arrest records are likewise often excluded from the data analysis and it would be rather easy to exclude data that might be serving as a proxy for discrimination. It would also be possible to correct algorithms after an experimental run, something along the lines of what occurred with Amazon. If it appears that the algorithm is producing discriminatory results, it can be altered to address that discrimination. This may not always be successful but between correcting a discriminatory algorithm and correcting human biases, again, the smart money should be on the algorithm.

All of this is to say that although algorithms certainly have the potential to produce discriminatory results, there is little reason to believe they will be more discriminatory than our existing systems and there are many reasons to believe they can be constructed to reduce discrimination. There remains the question of whether our existing legal structures can ferret out discrimination that arises through algorithms, the question to which we now turn.

III. ALGORITHMS AND THE LEGAL RESTRAINTS.

Antidiscrimination law is roughly divided into two broad spheres: (1) intentional discrimination and (2) unintentional discrimination, what is also known as the disparate impact theory.[100] Intentional discrimination is further divided into two classes, claims where intent is proved through traditional means and claims involving systemic discrimination that generally involve the use of statistics to prove intent.[101] The disparate impact theory seems the most likely means of challenging Type 2 algorithms, though there are also some scenarios where algorithms might be challenged under the systemic discrimination prong where intent is an element of the claim.[102]

As I demonstrated previously, disparate impact cases can be difficult to succeed on and most of the case law has developed around written employment examinations.[103] Existing law may prove to be an imperfect fit, but contrary to the concerns expressed by algorithm critics, it may actually prove difficult for employers or lenders to satisfy their burdens under that case law, and this will be particularly true for those Type 2 algorithms that are difficult for humans to decipher.

---

[100] *See* Michael Selmi, *Theorizing Systemic Disparate Treatment Law: After* Wal-Mart v. Dukes, 32 BERKELEY J. EMP. & LAB. L. 477, 478, 481–83 (2011).

[101] The two leading Supreme Court cases on systemic discrimination are *International Brotherhood of Teamsters v. United States*, 431 U.S. 324 (1977), and *Hazelwood School District v. United States*, 433 U.S. 299 (1977).

[102] *See infra* Part III.B.

[103] *See* Michael Selmi, *Was the Disparate Impact Theory a Mistake?*, 53 UCLA L. REV. 701, 702 (2006).

A number of commentators have suggested that algorithmic decisionmaking is likely to escape meaningful review under the disparate impact theory.[104] The concern, however, turns primarily on a single assumption, namely that a court will accept an argument, by an employer or lender, that while it does not know how the algorithm works, it knows that it does.[105] This concern obviously only relates to the black-box or Type 2 algorithm and even there, I believe the concern is overstated because allowing such an argument to succeed would effectively eliminate the possibility of demonstrating that there is an alternative practice that serves the employers needs with less adverse impact, an important but underused part of the disparate impact theory.[106]

Before proceeding to a discussion of systemic discrimination claims, let's first put aside potential claims of intentional discrimination, as it seems unlikely, though not implausible, that a biased programmer might intentionally create a discriminatory algorithm. This possibility does not seem consistent with what companies—employers and mortgage lenders in particular—are seeking to use algorithms for; in fact, algorithms are often employed as a way to diversify a workforce or to reach new borrowers.[107] To the extent a programmer engages in intentionally discriminatory practices through, for example, including various proxies for race or gender (zip codes perhaps), that algorithm would be subject to the same proof standards that exist for human decisions and should be no more or less difficult to prove. Again, I should reiterate, discrimination is never easy to prove but courts have been engaged in identifying intentional discrimination under Title VII for going on sixty years.[108] The only issue that may differ with algorithmic decisionmaking is those so-called black-box algorithms (Type 2) that are difficult to unravel, an issue where the disparate impact theory is most likely to be employed.

---

[104] *See* sources cited *supra* note 27; *see also* Ifeoma Ajunwa, *An Auditing Imperative for Automated Hiring Systems*, 34 HARV. J.L. & TECH. 621, 642–46 (2021) (expressing concern regarding disparate impact model); Pauline T. Kim, *Big Data and Artificial Intelligence: New Challenges for Workplace Equality*, 57 U. LOUISVILLE L. REV. 313, 326 (2019) (noting that "[e]xisting disparate impact doctrine is not equipped to deal with issues like [algorithms]").

[105] *See, e.g.*, Barocas & Selbst, *supra* note 18, at 706–09.

[106] *See* Albemarle Paper Co. v. Moody, 422 U.S. 405, 425 (1975) ("If an employer . . . meet[s] the burden of proving that its tests are 'job-related,' it remains open to the complaining party to show that other tests or selection devices, without a similarly undesirable racial effect, would also serve the employer's legitimate interest in 'efficient and trustworthy workmanship.'").

[107] *See* Cowgill & Tucker, *supra* note 90, at 22.

[108] The Court's decision in *McDonnell Douglas Corp. v. Green*, issued in 1973, remains the governing proof structure for individual claims of discrimination based on circumstantial evidence. *See* McDonnell Douglas Corp. v. Green 411 U.S. 792, 802–03 (1973). Similarly, two decisions issued in 1977 remain the primary authority regarding class claims of intentional discrimination. *See generally* Int'l Brotherhood of Teamsters v. United States, 431 U.S. 324 (1977); Hazelwood Sch. Dist. v. United States, 433 U.S. 299 (1977).

A. *The Disparate Impact Theory as Applied to Algorithms*

The identifying characteristic of the disparate impact theory is that it does not require proof of intent but instead a judicial inquiry is triggered when a selection device has a substantial disparate effect on a protected group.[109] If, for example, an employer's hiring practice leads to a predominantly white or male workplace, one that significantly differs from the group that applied, then under the theory, the law will require the employer to justify its practice.[110]

The Supreme Court established disparate impact liability as part of Title VII's statutory mandate back in 1971, and in 2015 the Court interpreted the Fair Housing Act to encompass disparate impact claims.[111] Although the disparate impact doctrine is far more developed under Title VII, claims under both statutes proceed in a three-part format: (1) the plaintiff must establish that an identified practice has caused a disparate effect; (2) the burden of proof then shifts to the defendant to justify its practice under what is known as a business necessity test; (3) if the defendant succeeds in justifying its practice, the plaintiff then has an opportunity to establish an alternative practice that would serve the defendant's needs with a less adverse impact that the defendant refuses to adopt.[112] This is, however, where challenges under the Constitution diverge given that the Supreme Court long ago held that proof of intentional discrimination is required and why some of the criminal cases are more complicated because of the unavailability of the disparate impact theory.[113]

1. *Establishing Disparate Impact*

In the first step of the proof process, the plaintiff has the burden (who has the burden will turn out to be significant) to identify a practice that has caused

---

[109] *See* Selmi, *supra* note 103, at 745.

[110] *See Albemarle*, 422 U.S. at 432–33 (discussing the business necessity test and what is required to satisfy that test).

[111] The Title VII case is *Griggs v. Duke Power Co.*, 401 U.S. 424 (1971), perhaps the most famous of Title VII cases. The standard was amplified in a case a few years later. *See generally Albemarle*, 422 U.S. 405. The Fair Housing case came much later, although the disparate impact theory had been recognized in lower courts for many years prior to the Supreme Court adoption of the theory. *See generally* Tex. Dep't of Hous. & Cmty. Affs. v. Inclusive Cmtys. Project, Inc., 135 S. Ct. 2507 (2015). The disparate impact theory is also available under the Age Discrimination in Employment Act ("ADEA") and the Americans with Disabilities Act ("ADA") but the standards vary, and under the ADEA in particular the standard is less rigorous than under Title VII. *See* Smith v. City of Jackson, 544 U.S. 228, 238–40 (2005) (ADEA establishing test of "reasonableness"). Disparate impact claims are less common under the ADA, in part because of the statutory accommodation requirement. *See* Bates v. United Parcel Serv., Inc., 511 F.3d 974, 995 (9th Cir. 2007) (en banc) (discussing disparate impact standard under the ADA).

[112] *Albemarle*, 422 U.S. at 435 (discussing alternative practice).

[113] Washington v. Davis, 426 U.S. 229, 238–39 (1976).

a statistically significant disparity based on one of the protected classes.[114] Under Title VII, this has most commonly been a written test, and zoning decisions have increasingly been the focus of disparate impact claims under the FHA.[115] This first step is purely statistical in nature, and for the most part should not cause legal concerns for challenges to algorithmic decisionmaking: the algorithm would be the practice that has disproportionately affected a protected group. The Amazon algorithm discussed previously—had it been used—would have likely had a statistically significant disparate effect on women, and at this first step that is really all that is necessary for a plaintiff to establish. A mortgage algorithm would be the same—the algorithm would adversely affect African-Americans or Latinos by denying their loan applications more frequently or charging them higher rates.[116] This is standard disparate impact fare and is readily adaptable to algorithmic decisionmaking.

A potential issue may arise if a court were to require the plaintiff to identify the particular part of the algorithm that is causing the disparate impact. Title VII requires a plaintiff to identify a particular practice that is causing the disparate impact, unless in the language of the statute, the "decisionmaking process [is] not capable of separation for analysis," in which case, "the decisionmaking process may be analyzed as one employment practice."[117] If anything qualifies as "not capable of separation," it would surely be a Type 2 algorithm, which by definition is inscrutable and therefore not capable of separation. When mortgage lending has been challenged, the lending process has been treated as a single

---

[114] *See* Ricci v. DeStefano, 557 U.S. 557, 587 (2009) (defining "a prima facie case of disparate-impact liability" as "essentially, a threshold showing of a significant statistical disparity"); Fudge v. City of Providence Fire Dept., 766 F.2d 650, 658 & n.8 (1st Cir. 1985) (holding that a prima facie case of disparate impact can be established where "statistical tests sufficiently diminish chance as a likely explanation").

[115] On Title VII, see, for example, *Johnson v. City of Memphis*, 770 F.3d 464 (6th Cir. 2014) (police promotional examinations); *Lewis v. City of Chicago*, 643 F.3d 201 (7th Cir. 2001) (fire department examination); *Bridgeport Guardians, Inc. v. City of Bridgeport*, 933 F.2d 1140 (2d Cir. 1991) (police exam); and *Griffin v. Carlin*, 755 F.2d 1516 (11th Cir. 1985) (post office promotional examination). On the Fair Housing Act, see, for example, *Summers v. City of Fitchburg*, 940 F.3d 133 (1st Cir. 2019) (zoning dispute); *Avenue 6E Investments, LLC v. City of Yuma*, 818 F.3d 493 (9th Cir. 2016) (zoning dispute); and *Reinhart v. Lincoln County*, 482 F.3d 1225 (10th Cir. 2007) (land use and zoning dispute).

[116] This is how mortgage lending cases have proceeded in the past when algorithms were not in use. *See, e.g.*, City of Los Angeles v. Wells Fargo & Co., 22 F. Supp. 3d 1047, 1051 (C.D. Cal. 2014) ("[A]n African-American borrower was more than twice as likely to receive a 'predatory loan' as a white borrower with similar underwriting and borrower characteristics."); Ramirez v. Greenpoint Mortg. Funding, Inc., 268 F.R.D. 627, 630 (N.D. Cal. 2010) (challenging "policy that led minority borrowers to be charged disproportionately high rates compared to similarly situated whites"). The Wells Fargo litigation was part of a series of complicated cases brought by municipalities seeking to remedy the harm that was done to their communities, and the Wells Fargo litigation was ultimately unsuccessful. *See* City of Los Angeles v. Wells Fargo & Co., 691 F. App'x 453, 454 (9th Cir. 2017).

[117] 42 U.S.C. § 2000e-2(k)(1)(B)(i).

practice and that will likely also be true for challenges to algorithms.[118] As a result, algorithms should not present any unique problems in the first step of the disparate impact analysis.

2. *Justifying the Practice*

It is the second step of the proof process where commentators have suggested the law is likely to come up short. In the second part of a disparate impact challenge, the burden of proof shifts to the defendant to justify its practice. Under Title VII, the justification must be "job related and consistent with business necessity" whereas the standard is a touch more vague under the FHA where the defendant needs to establish that the challenged policy is "necessary to achieve a valid interest."[119] A substantial amount of the case law under Title VII applies to written tests and specific standards have been developed to determine whether a test is valid, which typically requires proving that a test provides an employer with valuable information. One method of establishing validity that will likely be most relevant to algorithms is by demonstrating that there is a correlation between test scores and some measure of performance, often in the case of employment tests supervisor ratings of incumbent employees.[120] In this way, there is a statistical demonstration of whether those who perform well on the examination also perform well in the workplace. The correlation is rarely perfect (a perfect relationship would be 1.0, and most employment tests have correlations around .3) but courts almost always, with the aid of expert witnesses, can assess the quality of the statistical relationship.[121]

This takes us to the conundrum algorithm critics have identified.[122] By definition, an algorithm should have a statistically meaningful correlation to what it is trying to predict.[123] After all, the idea behind an algorithm is that it

---

[118] *See* Montgomery County v. Bank of Am. Corp., 421 F. Supp. 3d 170, 183 (D. Md. 2019); Prince George's County v. Wells Fargo & Co., 397 F. Supp. 3d 752, 766 (D. Md. 2019); County of Cook v. HSBC N. Am. Holdings, 314 F. Supp. 3d 950, 967 (N.D. Ill. 2010).

[119] Tex. Dep't of Hous. & Cmty. Affs. v. Inclusive Cmtys. Project, Inc., 135 S. Ct. 2507, 2522–23 (2015); *see also* Reyes v. Waples Mobile Home Park Ltd. P'ship, 903 F.3d 415, 419, 424 (4th Cir. 2018) (challenge to policy requiring tenants to demonstrate lawful legal status); Ellis v. City of Minneapolis, 860 F.3d 1106, 1112 (8th Cir. 2017) (challenge to housing enforcement policy).

[120] *See* Albemarle Paper Co. v. Moody, 422 U.S. 405, 431–32 (1975) (discussing validation effort correlating test scores with supervisor ratings); Ernst v. City of Chicago, 837 F.3d 788, 798 (7th Cir. 2016) (correlating test performance with performance ratings of paramedics); Lopez v. City of Lawrence, 823 F.3d 102, 119 (1st Cir. 2016) (incumbent employees took exam and correlated against performance ratings).

[121] *See, e.g.*, *Ernst*, 837 F.3d at 799.

[122] *See* Barocas & Selbst, *supra* note 18, at 706–09.

[123] Not all algorithms will produce correlations and absent a correlation, my sense is that an algorithm is not likely to be validated. *See id.*

will scour data to achieve a particular end goal—predicting musical tastes, borrower behavior, or productive employees. However, contrary to what some critics have suggested, establishing a meaningful correlation between the algorithm and employee performance or borrower behavior by itself would not be sufficient to establish business necessity under Title VII or validity under the FHA. Under existing law, it is not enough that an employer can establish a statistically significant correlation between a test and performance.[124] Indeed, in one of the early cases, the Supreme Court carefully reviewed an employer's justification for a written test even though it had established several statistically significant correlations between the exam and certain aspects of workplace skills.[125] Rather than accepting correlations as proof of business necessity, courts consistently assess whether the test also provides a basis for distinguishing among employees, which is particularly important when an employer intends to use a test to rank order the applicants for promotion or hiring.[126] Most tests, even when there is a statistical correlation between the test and performance, are simply not able to distinguish between candidates with modest point differences.[127] And when a practice has a disparate impact, a defendant will be required to justify the judgments the practice requires.

One reason it is important to analyze the underlying data has to do with the statistical principle of restriction of range. Although an algorithm might be able to select quality employees, it might not be able to determine whether other applicants might perform equally well.[128] This is effectively what is widely

---

[124] *Id.*

[125] *See Albemarle*, 422 U.S. at 432–33 (evaluating and rejecting employer's evidence despite the presence of statistically significant correlations on some measures). In a challenge to a lieutenant's test where the validity study included a correlation coefficient of .33, the court engaged in a lengthy analysis of the validity study before upholding the test, suggesting that courts do not simply accept a statistically significant correlation as proof of business necessity. *See* Hamer v. City of Atlanta, 872 F.2d 1521, 1527 (11th Cir. 1989).

[126] *See, e.g.*, *Ernst*, 837 F.3d at 804 ("We recognize that, in itself, there is nothing unfair about women characteristically obtaining lower physical-skills scores than men. But the law clearly requires that this difference in score must correlate with a difference in job performance."); El v. Se. Pa. Transp. Auth., 479 F.3d 232, 245 (3d Cir. 2007) ("[Title VII] require[s] that the policy under review accurately distinguish[es] between applicants that pose an unacceptable level of risk and those that do not."); Isabel v. City of Memphis, 404 F.3d 404, 414 (6th Cir. 2005) (concluding that there was "clear evidence that the scores from the written test did not approximate a candidate's potential job performance"); Bew v. City of Chicago, 252 F.3d 891, 895 (7th Cir. 2001) (quoting district court approvingly that the examination "must be scored so that it properly discriminates between those who can and cannot perform the job well").

[127] *See* Michael Selmi, *Testing for Equality: Merit, Efficiency, and the Affirmative Action Debate*, 42 UCLA L. REV. 1251, 1270–77 (1995).

[128] For discussions regarding the principle of restriction of range see generally John E. Hunter, Frank L. Schmidt & Huy Le, *Implications of Direct and Indirect Range Restriction for Meta-Analysis Methods and Findings*, 91 J. APPLIED PSYCH. 594 (2006), and Freddie deBoer, *Restriction of Range: What It Is and Why It Matters*, FREDDIE DEBOER BLOG (July 24, 2017), https://freddiedeboer.substack.com/p/restriction-of-range-what-it-is-and-why-it-

recognized as the LSAT problem: a law school might be able to determine that there is a correlation between LSAT scores and first year grades but without admitting applicants from a wider range of test scores, a school will not have evidence that students with lower test scores would not perform comparably well in school. This is a particular problem when correlations are modest, as is commonly the case not just with the LSAT but most selection procedures, including almost certainly most algorithms.[129] As a result, the applicants selected by the algorithm might prove to be valuable employees, but the algorithm cannot tell us whether those who were not selected would also have been good employees. This is something the law has always required and there is no reason to think that courts will now modify decades of law when confronted with a Type 2 algorithm.[130]

On the contrary, satisfying the existing legal standards will likely prove to be a particular problem for defendants to the extent the algorithm falls into the Type 2 category. In any disparate impact challenge, the defendant would be expected to explain what characteristics distinguish the selected from those who were not selected or those who obtained loans from those who did not. In the case of a Type 1 algorithm, the employer or lender will be able to identify differences that distinguish the two groups and the parties can then debate whether those differences justify the disparate results.[131] This is how litigation

---

matters [https://perma.cc/SAK3-8LLE]. There are ways to correct for a restricted range but given the nature of Type 2 algorithms, the corrections are not likely to be applicable. *See, e.g.*, Marie Wiberg & Anna Sundström, *A Comparison of Two Approaches to Correction of Restriction of Range in Correlation Analysis*, PRAC. ASSESSMENT, RSCH. & EVALUATION 1 (2009), https://scholarworks.umass.edu/pare/vol14/iss1/5/ [https://perma.cc/HF79-Y5T8].

[129] Alexia Brunet Marks & Scott A. Moss, *What Predicts Law Student Success? A Longitudinal Study Correlating Law Student Applicant Data and Law School Outcomes*, 13 J. EMPIRICAL LEGAL STUD. 205, 228 (2016).

[130] This is, in part, due to the role validation studies have played in the business necessity test. Under Title VII, many if not most litigated disparate impact claims have involved examinations, often in the context of police and fire departments, and the most common way to determine whether a test satisfies the business necessity standard is through a validation study. There are two common validation methods: one involves a test that is designed to measure or assess the content of the job, what is known as a content validation study. This seems likely to be unrelated to algorithms. The second validation method is known as a criterion-related study, one that seeks to establish a statistical and meaningful correlation between a predictor (the test) and some measure of performance, typically supervisor ratings in the context of a promotional examination. When a criterion-related study is used, there will typically be a means to determine whether those who performed well on the examination also performed well on the job and also to see how those who did not perform well on the examination did in the workplace. Validation studies are not required under the law but they have been common in Title VII disparate impact litigation. For some judicial discussions see *Hamer v. City of Atlanta*, 872 F.2d 1521, 1525–30 (11th Cir. 1989) (criterion-related validity), and *M.O.C.H.A. Society, Inc. v. City of Buffalo*, 689 F.3d 263, 281 (2d Cir. 2012) (content validity).

[131] This is how a typical employment disparate impact challenge proceeds. For example, in a gender discrimination challenge to a running test used to select transit officers, the question was whether running a mile and a half in twelve minutes was a reasonable criterion,

has proceeded under both statutes for decades. But for the defendant that is unable to articulate the basis for the selection method, their defense will likely come up short. After all, the core of a disparate impact claim is that the defendant's practice excludes members of a protected group in an unjustified way, and the only way to determine if it is unjustified is to know what the basis for the exclusion is.[132]

But, and this is an important point, if the algorithm demonstrates a statistically significant correlation with a valid measure of performance, then, under Title VII, the employer likely has satisfied the second step of the inquiry.[133] Similarly, if a lender can show those with large numbers of friends on Facebook are more likely to pay their loan on time than those who have fewer friends, that may suffice to establish a valid justification.[134] This assumes, of course, that the defendant is able to identify the quality that distinguishes the two groups, that in the language of one court, "distinguish[es] between applicants that pose an unacceptable level of risk and those that do not."[135] A defendant has to offer some means to compare those who are selected and those who are not, or those who obtained loans and those who did not.[136] By the same measure, if African-Americans or another protected group are more commonly on Instagram, then the lender should also be required to demonstrate the superiority of a Facebook network in terms of establishing creditworthiness. It is important to emphasize that there is nothing distinctive about an algorithm in this context, it is just a function of the legal standard that governs both Title VII and the FHA. Establishing the validity under the FHA or business necessity under Title VII is not, I should add, an easy test to meet, and it is not at all clear that all algorithms could satisfy the existing legal standards but substituting new for old criteria does not necessarily confound the legal analysis, it just requires an application to a new circumstance.

Consider mortgage lending. Assume a lender's algorithm has a disparate effect on African-Americans that leads to more denials on loan decisions and

---

which necessarily required analyzing whether those who obtained higher times would have also been more successful officers. This analysis was done by an expert witness when he established the time for the running test. *See* Lanning v. Se. Pa. Transp. Auth., 181 F.3d 478, 484, 489 (3d Cir. 1999).

[132] This is consistent with the requirement that employers can only test for minimum qualifications. *See* Isabel v. City of Memphis, 404 F.3d 404, 413 (6th Cir. 2005) (noting that cutoff scores should measure "minimal qualifications"); *Lanning*, 181 F.3d at 489 ("[A] discriminatory cutoff score [is impermissible unless shown to measure] the minimum qualifications necessary for successful performance of the job in question.").

[133] *See Lanning*, 181 F.3d at 486.

[134] Several legal scholars have expressed concern over an algorithm that apparently identified commuting distance as a key to employee longevity. *See* Kim, *supra* note 18, at 863. Although commuting distance could certainly have a disparate impact depending on where an employer was located, it is also the kind of criteria that an employer may be able to justify, particularly if employee turnover is an important concern.

[135] El v. Se. Pa. Transp. Auth., 479 F.3d 232, 245 (3d Cir. 2007).

[136] *Id.* at 239.

higher rates when loans are issued. The lender would likely contend that the differentials are due to higher risks, but it would also have to prove that there were, in fact, higher risks rather than simply asserting that justification.[137] And the risks would likely have to be meaningful rather than trivial. This is where mortgage lending even in the context of algorithms would effectively be analyzed much like a regression analysis, identifying the factors that went into the lending decision and assessing whether those factors might be treated differently for different groups.[138] Many existing studies have shown that African-Americans with similar credit profiles to whites still obtain inferior loan products,[139] what is the essence of discrimination, and if an algorithm produces similar discriminatory results, the legal analysis will not change.

A defendant is likely to succeed in its justification if it is able to identify distinguishing features relevant to the identified adverse impact. In the above example, if it turns out that African-Americans do pose greater credit risks, then the lender may be able to justify the higher rates it charges but again, those differences would have to be meaningful and based in data. The real problem for defendants will arise in those limited circumstances when an inscrutable algorithm is at issue; yet, contrary to what others have argued, this is more likely to be a problem for defendants than plaintiffs. If a Type 2 algorithm is responsible for the challenged decision, the defendant would likely be relegated to arguing that it does not know why the algorithm made the decisions it did, but it knows that it works and the reason it knows it works is because that is what it was designed to do. Sticking with the mortgage example, a lender would only be able to say that although it is lending disproportionately to white individuals, and it does not know why, it does know that the algorithm has determined this is the optimal means of measuring creditworthiness. This is akin to a plea to trust us, or more accurately, to trust the algorithm, an algorithm the defendant cannot explain.

---

[137] This principle was established in *Bazemore v. Friday* where the Court held that a defendant that is challenging a multiple-regression analysis must establish that its critique is steeped in facts rather than a theoretical concern. *See* 478 U.S. 385, 400 (1986) (reversing lower court's dismissal of plaintiffs' claim for failure to include "all measurable variables"). In the various residency cases, courts have required municipalities to come forward with a demonstrated need for residency rather than just asserting the potential benefits. *See, e.g.*, NAACP v. N. Hudson Reg'l Fire & Rescue, 665 F.3d 464, 480–82 (3d Cir. 2011) (rejecting city's claim that residency requirement made for quicker response time), *cert. denied*, 567 U.S. 906 (2012).

[138] Mortgage lending cases frequently involve regressions and also are frequently challenged under an intentional discrimination framework. *See, e.g.*, Ramirez v. Greenpoint Mortg. Funding, Inc., 268 F.R.D. 627, 632–34 (N.D. Cal. 2010); City of Los Angeles v. Wells Fargo & Co., 22 F. Supp. 3d 1047, 1051 (C.D. Cal. 2014) (using regressions in challenge to predatory lending). Regressions have likewise been common in many employment cases. *See, e.g.*, Moussouris v. Microsoft Corp., 311 F. Supp. 3d 1223, 1229–31 (W.D. Wash. 2018); Morgan v. United Parcel Serv. of Am., Inc., 380 F.3d 459, 466–72 (8th Cir. 2004).

[139] *See* studies cited *supra* note 116.

This is largely what the argument of the algorithm critics boils down to: would a court accept this "trust us" defense?[140] Based on existing case law, the answer should be a clear no—there is simply no precedent within the extensive disparate impact case law for judicial acceptance of such a defense. In no case has the defendant defended against a disparate impact challenge by arguing that even though we cannot explain our process, we know it works, and the reason we know it works is because that is what it was designed to do. Nor does it seem at all likely that a court would accept such a defense for at least two related reasons. The first and likely sufficient reason is that the defendant has the burden of proof on the question of the justification, and accepting a "trust me" defense would effectively insulate Type 2 algorithms from judicial review, thus emptying that phrase "burden of proof" of any content.

Although there are currently no cases involving inscrutable algorithms, there are analogous cases where courts have rejected what might be considered similar claims made by defendants. For example, in a case involving a running test for transit police officers in Philadelphia that had an adverse impact against female applicants, the employer (the very litigious SEPTA) argued that it should be allowed to establish a high cut score on a running test because being faster was always better.[141] The court, however, rejected the claim because it was inconsistent with the governing standards of business necessity and instead required the employer to show that the adopted standard was necessary to ensure applicants met the minimum standards.[142] Similarly, in an early case that challenged height and weight requirements for correctional officers, the Supreme Court rejected an argument that the requirements were related to strength, which, in turn, was relevant to being a correctional officer.[143] Although courts have frequently found tests or policies to be valid, in no case has a court simply deferred to an employer's judgment.[144] Nor have courts accepted

---

[140] This is the primary concern raised in the important article by Barocas & Selbst, *supra* note 18, at 709 ("[T]here is good reason to believe that any or all of the data mining models predicated on legitimately job-related traits pass muster under the business necessity defense."). In his self-styled "musing" on artificial intelligence, Professor Charles Sullivan has likewise noted, "[C]ourts have not previously been confronted with the argument that, however deficient a particular criterion seems to be, it can be empirically shown to be the best tool available." Sullivan, *supra* note 18, at 427; *see also* Selbst, *supra* note 27, at 1372 ("Because disparate impact doctrine ties legitimate employment criteria to statistical predictions of future outcomes, properly executed machine learning models will often pass muster.").

[141] Lanning v. Se. Pa. Transp. Auth., 181 F.3d 478, 484, 492 (3d Cir. 1999).

[142] *Id.* at 489.

[143] *See* Dothard v. Rawlinson, 433 U.S. 321, 331 (1977) (rejecting state's argument that height and weight requirements were related to strength because it failed to show "the requisite amount of strength thought essential to good job performance").

[144] As one court noted, "[a] business necessity standard that wholly defers to an employer's judgment as to what is desirable in an employee . . . is completely inadequate in combating covert discrimination based upon societal prejudices." *Lanning*, 181 F.3d at 490.

statistically significant correlations as proof of business necessity without further inquiry.[145]

As these cases illustrate, the disparate impact theory demands trade-offs. An employer or lender is free to structure its practices any way that it likes so long as it does not have a disparate impact, but once that impact is shown, the defendant must justify its practice as necessary to the business.[146] Preferred and necessary are not equivalents, and the necessary justification always involves comparisons—can the person who runs a bit slower still function effectively as a transit officer? Can the person who scores a point lower on a written examination still be a successful lieutenant in the fire department? Rarely will a single criterion—a test or an algorithm—predict performance perfectly, in which case it is necessary to weigh the value of the criterion against its disparate impact. Under the law going back now fifty years, an employer cannot defend against a disparate impact by arguing that the challenged practice is what is best for the business; the law has determined that what is best for society is to balance a defendant's own interests with a societal desire to limit discrimination, whether intentional or not.[147]

It is true, as some critics have noted, that courts have, on occasion, deferred to well-constructed professional examinations, and it is possible that a court would be persuaded about the power of the algorithm simply by the way it was constructed.[148] In the Amazon example discussed earlier, Amazon might explain to the court that it introduced the algorithm to reams of data regarding its existing workforce, including most likely performance evaluations, and the algorithm set out to identify characteristics that were indicative of success on

---

[145] For a sampling of cases where courts rejected correlations, typically associated with criterion-related studies, see *Bernard v. Gulf Oil Corp.*, 841 F.2d 547, 567 (5th Cir. 1988) (remanding case despite correlations in range of .22 to .51), and *Arndt v. City of Colorado Springs*, 263 F. Supp. 3d 1071, 1082 (D. Colo. 2017) (rejecting validity of physical agility test because correlations were low and cut-off scores arbitrary). *See also* cases cited *supra* notes 125–26.

[146] *See Lanning*, 181 F.3d at 490.

[147] *See* Griggs v. Duke Power Co., 401 U.S. 424, 430–32 (1971).

[148] To the extent courts have deferred to professionally developed examinations, it has been in the context of public safety jobs. *See* Johnson v. City of Memphis, 770 F.3d 464, 478 (6th Cir. 2014) ("When the employment position involves public safety, we accord greater latitude to the employer's showing of job-relatedness and business necessity."). A number of commentators have also raised a concern that courts will defer to algorithms because they appear to be scientific and objective. *See* EUBANKS, *supra* note 20, at 179 (expressing concern that algorithms are perceived as infallible); BENJAMIN, *supra* note 20, at 53 (stating that algorithms have "the allure of objectivity without public accountability"). This is certainly a possibility, but courts frequently deal with technical issues and to date have had little trouble analyzing algorithms in various contexts. *See* Hous. Fed'n of Tchrs., Loc. 2415 v. Hous. Indep. Sch. Dist., 251 F. Supp. 3d 1168, 1180 (S.D. Tex. 2017) (concluding that the use of a proprietary algorithm to determine terminations violated teacher's due process); Chacko v. Connecticut, No. 3:07-cv-1120, 2010 WL 1330861, at *8–9 (D. Conn. Mar. 30, 2010) (allowing discrimination challenge to algorithm that assigned work in a hospital to survive summary judgment).

the job. In analyzing the data, the algorithm determined that being a man was a central criterion. But just stating this argument reveals why it is unlikely to prevail. Surely, at some point in the litigation, Amazon would be asked to explain why gender was the important criteria, and they would presumably be unable to do so. Given the reported facts, they would likely also be asked why women were penalized for having female indicators on their resumes, and again, Amazon would be able to do little more than shrug its shoulders and possibly add that the algorithm determined it was relevant. It is difficult to see a court accepting such an explanation, particularly when the algorithm may well have been created on biased data, namely that there were simply few female employees rather than female employees were less competent than their male counterparts. On the contrary, as discussed in the next Part, consciously building an algorithm on biased data would likely be challenged under a theory of intentional discrimination rather than under the disparate impact theory.[149]

The second reason courts are unlikely to accept a simple correlation to justify an algorithm is directly related to this last point and is perhaps the most compelling reason the "trust us" defense should fail. The third step of the disparate impact proof structure allows a plaintiff to prove that there is an alternative practice that would serve the defendant's interests while reducing adverse impact.[150] There is also a safety valve for defendants that allows them to avoid liability by adopting the proposed alternative, at least under Title VII and likely to be imported in to the FHA as well.[151] But this third prong can only work if the plaintiff is able to understand the nature of the algorithm and how it makes its decisions.

3. *Establishing a Lesser Discriminatory Alternative*

This third prong of the disparate impact theory has generally been underdeveloped in the case law, but it is likely to play a significant role with respect to algorithmic decisionmaking because altering the algorithm may reduce discrimination without significantly affecting the quality of the decisions. This has been true in a number of testing cases where plaintiffs propose an alteration of the scoring regime, for example by moving away from rank order selection to a broader scoring system, and the same could easily be done with mortgage lending where certain inputs might be weighted differently.[152]

---

[149] *See infra* Part III.A.3.

[150] *See* Albemarle Paper Co. v. Moody, 422 U.S. 405, 425 (1975) ("If an employer does . . . meet the burden of proving that its tests are 'job related,' it remains open to the complaining party to show that other tests or selection devices, without a similarly undesirable racial effect, would also serve the employer's legitimate interests in 'efficient and trustworthy workmanship.'").

[151] *See* 42 U.S.C. § 2000e-2(k)(1)(A)(ii) (Title VII).

[152] *See, e.g.*, Allen v. City of Chicago, 351 F.3d 306, 311–13 (7th Cir. 2003) (proposing change in process for merit promotions as an alternative); Bridgeport Guardians, Inc. v. City

But this third prong is only viable if the algorithm can be revealed and analyzed. If a defendant were allowed to prevail under a general claim that the algorithm is inherently accurate though we do not know why, it would be impossible for a plaintiff to offer any alternative. As a result, not only would such a defense stretch the concept of the burden of proof, but it would eliminate the third step in the analysis, a step that has been enshrined in disparate impact law for nearly fifty years and now safely ensconced in the statutes.

This also creates an interesting legal paradox that has largely escaped notice: assuming that a court will not accept a defense based solely on a statistically significant correlation between the algorithm and some meaningful measure of performance, this likely means that a defendant could not survive a challenge to a black-box algorithm. It is possible that a defendant would be able to offer alternative algorithms to show the superiority of its own, though unless one knows the underlying construct it would be hard to compare the algorithms and equally difficult to know whether the alternative was used solely to demonstrate the merits of the company's algorithm rather than as a true considered alternative.[153] As more work is done on algorithmic equity, these issues will likely be overcome or refined, but until then, a challenge to an inscrutable algorithm that has disparate impact is likely to mean either that the plaintiffs will always or generally prevail or the defendants will, and under the terms of the statutes, particularly with the focus on alternatives, if such a choice needs to be made, it should be in favor of the plaintiffs, otherwise the defendants' practices would be effectively insulated from review.

B. *The Pattern or Practice Theory*

There is one other argument that may be appropriate in certain challenges to algorithms, and one that has the potential for substantially better remedies. In addition to the two theories I have already discussed—classic disparate treatment and disparate impact—there is a third theory that involves class action intentional discrimination proved through statistical analysis.[154] This theory is known as a pattern and practice claim but is also frequently labelled as involving systemic discrimination.[155] Shortly after the Civil Rights Act of 1991 was enacted, there was a surge of pattern and practice claims that has since been

---

of Bridgeport, 933 F.2d 1140, 1148 (2d Cir. 1991) (proposing alternative involving a band of scores); *see also* Jones v. City of Boston, 845 F.3d 28, 32–34 (1st Cir. 2016) (proposing alternative drug testing process to replace hair testing).

[153] A plaintiff potentially could employ their own machine learning to try to devise an algorithm with less adverse impact, and this might be one way of searching for substantially equivalent alternatives but at the same time this would seem to impose a burden, particularly financially, that exceeds what the law intended as identifying alternative practices.

[154] *See generally* Int'l Brotherhood of Teamsters v. United States, 431 U.S. 324 (1977); Hazelwood Sch. Dist. v. United States, 433 U.S. 299 (1977).

[155] *See, e.g.*, Michael J. Zimmer, Charles A. Sullivan & Rebecca Hanner White, *Taking on an Industry: Women and Directing in Hollywood*, 20 EMP. RTS. & EMP. POL'Y J. 229, 262–71 (2016) (discussing systemic discrimination litigation).

tempered somewhat by the Supreme Court's *Wal-Mart Stores, Inc. v. Dukes* decision that was widely perceived to have made class actions more difficult to certify.[156] But the *Wal-Mart* case only indirectly addressed the legal standard governing pattern and practice claims, which remain viable and potentially quite potent.[157]

In many ways, the pattern and practice claims mirror the proof process for disparate impact claims, with three important differences. Although the statistical proof under both theories is largely the same—a plaintiff must establish a statically significant disparity against a protected group—in a pattern or practice claim, the statistics are used to prove intent, to prove in the words of an influential older case, that discrimination was the company's "standard operating procedure."[158] Relatedly, because the case is based on a theory of intentional discrimination, damages are available whereas under a disparate impact theory only lost wages and injunctive relief are available.[159]

Another difference and a key aspect of the pattern-or-practice claim is that although the cases are steeped in intentional discrimination, it is not necessary to demonstrate that the defendant adopted the practice so as to exclude women or minority borrowers. Rather the statistical proof provides evidence of intent.[160] Significantly, and unlike the disparate impact theory, there is also no business necessity defense available for a claim of intentional discrimination, an employer is not able to justify its practice as valuable, efficient, or rational. Rather an employer, or a lender, is limited to challenging the plaintiff's statistical proof, which may include providing some alternative explanation to avoid drawing an inference of discrimination.[161] This area of the law is relatively undeveloped, but it tends to focus on statistical battles over how to interpret the data.[162] For example, in the *Wal-Mart* sex discrimination case, the

---

[156] *See generally* Wal-Mart Stores, Inc. v. Dukes, 564 U.S. 338 (2011).

[157] I have previously described the pattern-or-practice claim as "the most potent but least understood of the various Title VII causes of action." Selmi, *supra* note 100, at 478.

[158] *Teamsters*, 431 U.S. at 336.

[159] *See id.* at 382. The damages are capped at a maximum of $300,000 per plaintiff and as a result, the damages can be substantial depending on the size of the class. *See* 42 U.S.C. § 1981a(b)(3)(D).

[160] *Id.* at 337–39 (relying on statistical and anecdotal evidence in determining that the government carried its burden of proof in establishing a prima facie case of pattern or practice discrimination).

[161] In one of the more well-known claims of systemic discrimination, the plaintiffs established that women were largely excluded from the best paying jobs at Sears, to which Sears successfully argued that women lacked interest in the jobs, many of which paid on commission. *See* EEOC v. Sears, Roebuck & Co., 839 F.2d 302, 313 (7th Cir. 1988). In the case, the defendants challenged the meaning of the statistics rather than offering a justification for its practice. *Id.* at 313–14. For a more recent case involving salary disparities see *Velez v. Novartis Pharmaceuticals Corp.*, 244 F.R.D. 243 (S.D.N.Y. 2007).

[162] *See, e.g.*, Hazelwood Sch. Dist. v. United States, 433 U.S. 299, 306 (1977) ("[P]etitioners primarily attack[ed] the judgment of the Court of Appeals for its reliance on 'undifferentiated work force statistics to find an unrebutted prima facie case of employment discrimination.'"); *Sears, Roebuck & Co.*, 839 F.2d at 312 ("[M]ost of Sears' evidence was

company argued that decisions should be analyzed at the local—as opposed to the national—level, and in the mortgage lending cases, the lenders typically argue that there are many players such as independent mortgage brokers that are involved and effectively break up the causal chain.[163]

Although most analyses to date have focused on the disparate impact theory, mortgage lending cases have often been brought as pattern or practice claims and the theory seems particularly appropriate when the defendant consciously relies on data that is known to be biased.[164] This likely would have been true in the Amazon example.

C. *Altering the Algorithms*

It is easy to forget that a primary interest in algorithmic decisionmaking was not just to create better decisions but also to reduce the discrimination that continues to affect so many human decisions. Objective data, it was hoped, would remove what is often defined as subtle or implicit bias and would root out any effort to engage in more intentional schemes.[165] But we have quickly learned that the data are not so objective and discrimination lurks behind or within many algorithms and the data they are built on.[166] As a result, and because it is assumed that many employers or lenders would only want to use algorithms if they were nondiscriminatory, there has been considerable legal attention paid to whether an employer might be able to alter its algorithm if it discovers that it has adverse impact.[167]

Amazon never actually used the algorithm to hire anyone but if it had, and it found out after it implemented the algorithm that men were overrepresented in the group that was hired, could the company alter its algorithm moving forward, or could it alter it even after its initial implementation as a way of changing who was hired? These manipulations obviously assume that Amazon has access to its own algorithm, including the weights various factors are

---

directed at undermining two assumptions Sears claimed were faulty and fatal to the validity of the EEOC's statistical analysis . . . .").

[163] *See, e.g.*, Prince George's County v. Wells Fargo & Co., 397 F. Supp. 3d 752, 758, 766 (D. Md. 2019) (rejecting defendant's argument that it was not responsible for all of the various steps in the process).

[164] Most of the wave of cases brought by local governments challenging the lending practices of banks included pattern and practice claims. *See, e.g.*, Montgomery County v. Bank of Am. Corp., 421 F. Supp. 3d 170, 178 (D. Md. 2019); County of Cook v. Wells Fargo & Co., 314 F. Supp. 3d 975, 995–96 (N.D. Ill. 2018); City of Los Angeles v. Citigroup Inc., 24 F. Supp. 3d 940, 952–54 (C.D. Cal. 2014); Nat'l Fair Hous. All., Inc. v. HHHunt Corp., 919 F. Supp. 2d 712, 715 n.1 (W.D. Va. 2013). Courts have long borrowed Title VII's pattern and practice theory in Fair Housing Act cases. *See* Gamble v. City of Escondido, 104 F.3d 300, 305 (9th Cir. 1997).

[165] *See* Sullivan, *supra* note 18, at 399–400.

[166] *See* Barocas & Selbst, *supra* note 18, at 673–74.

[167] *See id.* at 725–26; Pauline T. Kim, *Auditing Algorithms for Discrimination*, 166 U. PA. L. REV. ONLINE 189, 191–93 (2017); Kroll et al., *supra* note 18, at 694–95.

afforded, otherwise it would run into the same problem that plaintiffs would have in proposing an alternative to a black-box algorithm. A party could likely only alter an algorithm that it could understand, though it would clearly be possible to change the underlying data after seeing the results even if the algorithm itself proved inscrutable.

One of the reasons this issue has received so much attention is due to a Supreme Court case that appears to be squarely on point, though there are some important differences.[168] The case—*Ricci v. DeStefano*—involved promotional tests administered by the City of New Haven for various supervisor positions in its fire department.[169] After administering the tests, it was clear based on the civil service provisions governing fire department promotions that nearly all of the promotions would go to white men.[170] This was in a diverse city with a diverse group of firefighters in a department with a long history of discrimination.[171] As a result of the disparate impact, the city effectively discarded the test results by not certifying them, and they were then sued by a group of white and Latino firefighters who likely would have been promoted over the two-year life of the list based on their test scores.[172]

The legal challenge was brought under the Equal Protection Clause and had some semblance of an affirmative action case in that the plaintiffs were challenging the fire department's race conscious action that was designed to mitigate the examination's disparate impact.[173] And that semblance or overlay led to the Court's determination that: "We conclude that race-based action like the City's in this case is impermissible under Title VII unless the employer can demonstrate a strong basis in evidence that, had it not taken the action, it would have been liable under the disparate-impact statute."[174] The altering of an algorithm to achieve different results from what was originally produced feels similar to discarding test results, which leads to the question what might constitute a "strong basis in evidence" that would justify altering an algorithm.[175]

Judge Calabresi has provided the most extensive analysis of what the *Ricci* Court intended and several subsequent Second Circuit cases have expanded on

---

[168] Ricci v. DeStefano, 557 U.S. 557 (2009).

[169] *Id.* at 562.

[170] *Id.*

[171] In her dissenting opinion, Justice Ginsburg highlighted the fire department's history of discrimination including litigation that commenced in the 1970s. *See id.* at 608–11 (Ginsburg, J., dissenting).

[172] *Id.* at 562–63 (majority opinion).

[173] *Id.* at 563.

[174] *Ricci*, 557 U.S. at 563.

[175] In *Ricci*, the Court borrowed the standard from its affirmative action doctrine, though the phrase remains relatively undertheorized. *See id.* at 582 (citing Richmond v. J.A. Croson Co., 488 U.S. 469, 500 (1989)).

that analysis.[176] Writing for the court in *United States v. Brennan*, issued shortly after *Ricci* and curiously ignored by the algorithm literature, Judge Calabresi noted that a strong basis in evidence requires more "than speculation" and "more than a mere fear of litigation, but less than the preponderance of the evidence that would be necessary for actual liability."[177] It does not, in other words, require the employer (or lender) to prove that it would lose any litigation but rather there has to be "an objectively reasonable basis to fear such liability."[178] This might include evidence of clear disparate impact along with some objective evidence that the practice may not be justified under the business necessity test.[179] This is not an easy standard to meet, and ultimately the City of New Haven failed to do so in complicated follow-up litigation.[180] But in at least one case, the Second Circuit upheld the right of the City of Buffalo to use a new test rather than one with demonstrated adverse impact, even when the city was motivated by a desire to decrease the racial disparity in the prior test results.[181]

How the "strong basis in evidence" standard will play out in the case of altering algorithms to produce different results will likely depend on the context, including whether the entity is a public or private actor. The worst case scenario is likely that which was present in *Ricci*, where the individuals largely knew their place on the list and had what might be described as reasonable expectations of being promoted based on the test results.[182] As Pauline Kim has recently noted, without that expectation, the situation would have been decidedly different,[183] as evidenced by the fact that the City of Buffalo was able to move to a new test rather than relying on its past exam.[184] Moreover, in *Ricci*, the city was bound contractually to count the written examination as 60% of the

---

[176] *See, e.g.*, United States v. Brennan, 650 F.3d 65, 109–14 (2d Cir. 2011); Briscoe v. City of New Haven, 654 F.3d 200, 205–09 (2d Cir. 2011); Maraschiello v. City of Buffalo Police Dep't, 709 F.3d 87, 95 (2d Cir. 2013).

[177] *Brennan*, 650 F.3d at 109–10.

[178] *Id.* at 113.

[179] *Id.* at 109. Ironically, a poorly constructed test may be the best hedge against discriminatory results.

[180] Following the *Ricci* decision, African-American firefighters sued to challenge the test as having an unjustified adverse effect. The Second Circuit allowed the case to move forward but the disparate impact claim was ultimately dismissed by the District Court. *See Briscoe*, 654 F.3d at 209–10, *remanded* 967 F. Supp. 2d 563 (D. Conn. 2013).

[181] *Maraschiello*, 709 F.3d at 95–96.

[182] In a strange move that may have exacerbated tensions, the city only published the race of the test takers, not the names, so people knew that, for example, a white individual placed in certain position, as the test was designed to be used for rank-ordered hiring. Ricci v. DeStefano, 557 U.S. 557, 567 (2009). I have previously written about the *Ricci* case and at an American Association of Law Schools meeting appeared on a panel with, among others, the City Attorney for New Haven. *See* Michael Selmi, *Understanding Discrimination in a "Post-Racial" World*, 32 CARDOZO L. REV. 833, 844–54 (2011).

[183] Kim, *supra* note 167, at 199.

[184] *See Maraschiello*, 709 F.3d at 90.

total and the oral portion 40%, so it was not possible to alter that mix, at least without being sued for breach of contract.[185]

For an employer or entity concerned about the potential disparate impact of its practice, the best approach would be to try out the practice on some group, most likely incumbent employees but in the case of a machine learning algorithm, it could also be tested on a potential group of hires. Surely no one would think what Amazon did—create an algorithm that it never uses for hiring purposes—was somehow discriminatory, so testing out a practice should be entirely lawful, and the process I just mentioned seems akin to what the company Pymetrics has done in creating a hiring game for the law firm O'Melveny where the company pledged to ensure the game does not have adverse impact.[186] This is also a significant advantage to algorithms over written examinations, as it is relatively easy to run different analyses based on varied data whereas it would be nearly impossible (and administratively cumbersome) to have individuals retake a test on multiple occasions. It should be noted that there would be no guarantee that a selection practice that did not have a disparate impact on a test group would also not have an adverse impact on some other group such as actual applicants. There may be ways to assess this likelihood[187] but it is entirely possible that a practice that was thought to be free of bias in one algorithmic application would later have a significant adverse effect on another. At the same time, testing out an algorithm before it is used would surely be permissible and may provide a means to minimize or even eliminate adverse impact.

Given the previous discussion, the use of an algorithm that has disparate impact may also have a relatively easy time satisfying *Ricci*'s "strong basis in evidence" standard.[188] Assuming a defendant will be unable to justify its algorithm under the business necessity test, then it should have the freedom to alter its algorithm because it would have a clear indication that it would likely lose any litigation regarding the use of the algorithm.

Private employers are also likely to have greater leeway to adjust algorithms, in part because they are not constrained by civil service rules or the Equal Protection Clause and will likely have greater flexibility in structuring

---

[185] *Ricci*, 557 U.S. at 589.

[186] O'Melveny garnered considerable attention when it moved to a game process for hiring outside of its traditional recruitment process. The company that designed the game has publicly ensured that its results are nondiscriminatory. *See* Victoria Hudgins, *Diversity, Metrics Demands Are Pushing Firms to Embrace AI Hiring Tools*, LEGALTECH NEWS (Jan. 13, 2021), https://www.law.com/legaltechnews/2021/01/13/diversity-metrics-demands-are-pushing-firms-to-embrace-ai-hiring-tools/ (on file with the *Ohio State Law Journal*).

[187] There is considerable research regarding how one might best ensure unbiased algorithms. *See, e.g.*, Michael Veale & Reuben Binns, *Fairer Machine Learning in the Real World: Mitigating Discrimination Without Collecting Sensitive Data*, BIG DATA & SOC'Y (Nov. 20, 2017), https://journals.sagepub.com/doi/pdf/10.1177/2053951717743530 [https://perma.cc/T9FM-L2VA]; Kim, *supra* note 167.

[188] *Ricci*, 557 U.S. at 582 (citing City of Richmond v. J.A. Croson Co., 488 U.S. 469, 500 (1989)).

their hiring practices.[189] But it may depend on how they go about addressing disparate impact. If a selection process, like the one Amazon developed, has an adverse impact on women, some employers might be tempted to go outside the algorithm to hire more women to balance out the effect of the algorithm. Doing so, however, would likely run afoul of a principle established by the Supreme Court many years ago, namely that having a "bottom-line" that is unbiased is not a defense to a biased employment practice.[190] Moreover, the overt use of race or gender to address an algorithm's disparate effect is likely to be seen as a form of intentional discrimination that might only be justified in the manner prescribed by *Ricci*.

Private employers, however, are also likely to have greater flexibility for another reason—most employees would never know how the employer made its decisions. Using the affirmative action analogy, private employers are rarely sued for their voluntary affirmative action programs, and when they are sued, they generally settle the cases without litigation. In his excellent book *After Civil Rights*, Professor John Skrentny documents how employers frequently rely on various forms of affirmative action that would almost certainly be statutorily suspect and yet they are rarely, if ever, called on it.[191] As a result, private employers would likely be able to adjust their algorithms to reduce or eliminate without substantial fear of legal liability.

D. *Algorithms and Trade Secrets*

One final issue that a number of commentators have suggested may pose problems for evaluating algorithms is the fact that many private companies treat their algorithms as a trade secret and may resist revealing even Type 1 algorithms for fear that disclosure might lead to the loss of trade secret protection.[192] In the context of litigation, this is certainly a non-issue. Indeed, if a party could stymie litigation by asserting trade secret protection, there would be no litigation regarding trade secrets, including their infringement. In a case involving an assertion of trade secret protection, the proper approach is for a court to issue a protective order limiting those who can have access to the trade secret and limiting the use of any information obtained during the litigation.[193] It is also quite likely that if challenged in court many asserted claims of trade

---

[189] Kim, *supra* note 167, at 199.

[190] Connecticut v. Teal, 457 U.S. 440, 452 (1982).

[191] *See generally* JOHN D. SKRENTNY, AFTER CIVIL RIGHTS: RACIAL REALISM IN THE NEW AMERICAN WORKPLACE (2014).

[192] *See, e.g.*, Kim, *supra* note 18, at 921 (noting that "the algorithm's creators are likely to claim that both the training data and the algorithm itself are proprietary information").

[193] *See, e.g.*, *In re* City of New York, 607 F.3d 923, 935 (2d Cir. 2010) ("The disclosure of confidential information on an 'attorneys' eyes only' basis is a routine feature of civil litigation involving trade secrets."); Sioux Pharm, Inc. v. Eagle Lab'ys, Inc., 865 N.W.2d 528, 537–41 (Iowa 2015) (discussing provisions for guarding trade secrets). The issue is discussed fully in Rebecca Wexler, *Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System*, 70 STAN. L. REV. 1343, 1409–34 (2018).

secret for algorithms would fail, particularly if the algorithm's code could be reverse engineered so as to be "readily ascertainable."[194]

IV. CONCLUSION

The novelty and perceived objectivity of algorithms have raised new and important questions for how the legal system will analyze the algorithms to ensure consistency with existing antidiscrimination principles. These questions become easier to address once it is clear that the vast majority of algorithms in use are not the so-called black-box algorithms but are instead complicated procedures with identifiable parts that largely resemble complex statistical models, like regressions, that have been analyzed for decades. And contrary to the concerns raised by many critics, the black-box algorithms, what I have referred to as a Type 2 algorithm, are likely to pose more problems for defendants than plaintiffs under existing case law, which largely mutes the concerns among the critics that algorithmic decisionmaking will escape meaningful judicial review.

[194] Wexler, *supra* note 193, at 1413–14.