Volume 63 | Issue 3 Article 2

10-1-2018

Employing AI

Charles A. Sullivan

EMPLOYING AI

CHARLES A. SULLIVAN*

I. INTRODUCTION

THOUGHT experiments can be useful not only in exploring new concepts but also in bringing interesting perspectives to bear on old problems. Developments in "people analytics," perhaps someday leading to the dominance of artificial intelligence in selecting and managing employees, offer an opportunity to do both. One disturbing conclusion from such an examination is that current paradigms do not seem to reach even the explicit use of race, sex, or other "protected classes" as selection criteria when deployed by artificial intelligence. That has important implications for current law, entirely apart from actual developments in AI. In addition, such an examination sheds new light on how other aspects of Title VII may apply to the spread of Big Data to the workplace.

* * *

Imagine that, today or in the not-so-distant-future, a company desires to take full advantage of the developments of artificial intelligence by effectively delegating all its hiring decisions to a computer.[1] It gives the computer only one instruction: "Pick good employees." Taking "Big Data" to the logical extreme, the computer (we'll call it Arti[2]) is also provided with all the employer's available data. That includes not merely what

* B.A., Siena College; LL.B, Harvard Law School. Professor of Law and Senior Associate Dean for Finance & Faculty, Seton Hall Law School. This article draws in part from my posting on JOTWELL, *Comprehending Causation and Correlation*, Aug. 4, 2017, https://worklaw.jotwell.com/#identifier_0_1087 [https://perma.cc/4ZYY-RB5P]. It also draws from various thoughts expressed in SULLIVAN & ZIMMER, CASES & MATERIALS ON EMPLOYMENT DISCRIMINATION (2017). I thank Neil Cohen, Timothy Glynn, Ed Hartnett, Najarian Peters, Jon Romberg, Rebecca Hanner White, Steve Willborn, and Elana Zeide for indulging my ruminations and correcting my more egregious errors. Also, thanks to Charles Mueller, Seton Hall class of '19, who provided able research assistance.

1. While this article imagines a scenario well beyond what current technology can achieve, "people analytics" is making inroads in the workplace today, and the trend seems certain to increase. *See, e.g.*, Pauline Kim, *Data-Driven Discrimination at Work*, 58 WM. & MARY L. REV. 857, 860 (2017) [hereinafter Kim] (documenting some current efforts to create algorithms for employee selection); *see also* Jennifer Alsever, *How AI is Changing Your Job Hunt*, FORTUNE (May 19, 2017), http://fortune.com/2017/05/19/ai-changing-jobs-hiring-recruiting/ [https://perma.cc/4WZ2-SABX]. Unlike the scenario imagined here, however, human beings are heavily involved in current data mining in the employment arena. For a more general critique of the effects of big data on society, see FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION (2015).

2. "Hal" was taken. *See* 2001: A SPACE ODYSSEY (Metro-Goldwyn-Mayer 1968). Hal stands for **H**euristically programmed **AL**gorithmic computer.

might be termed "human resources" information about its workers (such as performance evaluations) but also data on all the employer's operations because individual performances might be linked to particular successes and failures of the business. To cap it off, the company authorizes the computer to scour the internet to identify other traits of successful and unsuccessful current and former employees in order to assist its task.[3] Although there are obvious privacy concerns with such an exercise,[4] my focus is on implications under the antidiscrimination laws.[5]

The instruction "pick good employees," to say the least, provides hardly any guidance,[6] but at the 30,000-foot level is the goal of any employer. Our hypothetical firm believes Arti will be competent at its job; that is, it expects it to select persons who turn out to be good employees, or at least do better at that task than humans who lack both its processing power and its data resources and therefore will outperform Arti only by luck.[7]

---

3. It might, in addition, develop data not currently available. For example, machines can apparently conduct first-round interviews with applicants in which the goal is to obtain information about their qualifications not available from a traditional application or resume. *See* Simon Chandler, *The AI Chatbot Will Hire You Now*, WIRED.COM (Sep. 13, 2017), https://www.wired.com/story/the-ai-chatbot-will-hire-you-now/ [https://perma.cc/3YCL-BLYY]; *see also* Matthew T. Bodie, Miriam A. Cherry, Marcia McCormack & Jintong Tang, *The Law and Policy of People Analytics,* 88 U. COLO. L. REV. 961, 973-85 (2017) [hereinafter Bodie] (reporting how some information is gathered by requiring applicants to play games supposedly revealing information about their talents and skills).

4. *See* Bodie, *supra* note 3, at 985-1007.

5. The two concerns, however, overlap, most prominently with respect to federal laws such as the Americans with Disabilities Act, 42 U.S.C. § 12112(d) (2018), and the Genetic Information Nondiscrimination Act, 42 U.S.C. § 2000ff-1 (2018), which bar employer "inquiries" about medical or genetic conditions in order to discourage discrimination. Whether Arti's mere access to data that could reveal disabilities or genetic information would violate those statute's prohibitions on employer inquiries is another question, the answer to which might depend on whether Arti sought out new information or contented itself with what the employer had currently available. Data gathering could also uncover protected class membership for applicants, such as sex or race, which might not otherwise be available. One example of analytics uncovering such information, albeit in the consumer context, is Target's developing customer profiles, which generated information about likely pregnancy. Bodie, *supra* note 3, at 999. Such information could obviously be used to discriminate.

6. *See* Kim, *supra* note 1, at 875 (suggesting that data analysis begins with a contestable definition of what constitutes a "good" employee); Solon Barocas & Andrew Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 679 (2016) [hereinafter Barocas & Selbst] ("[T]he definition of a good employee is not a given. 'Good' must be defined in ways that correspond to measurable outcomes: relatively higher sales, shorter production time, or longer tenure, for example."). In Arti's case, however, that decision as to who counts as a good employee is outsourced to it to begin with.

7. Admittedly, the employer might have to provide some further instruction to the computer regarding the goals of the company: a "good" employee will assist a firm in achieving its goals, and different goals might require different traits for good employees.

To carry out its mission, Arti will need to define what makes an employee "good," and might well develop different criteria for different positions. So a good worker in a physically demanding job might be one who is physically strong while a good worker in a research lab might have a Ph.D. in one of the sciences. However it goes about its task, there are two obvious constraints on Arti's ability to define a "good" employee, both stemming from the data it has available.

The first constraint is defects in the data. For example, individual pieces of information may be incorrect or correct information may be aggregated to yield inaccurate results, as when two individuals with similar names are treated as the same person. Alternatively, the data may be accurate as far as it goes but problematic insofar as it incorporates prior discrimination, as when past performance evaluations are tainted by conscious or unconscious bias.[8] The data may also be unrepresentative or incomplete. Thus, it may be skewed if members of protected classes are not adequately represented. For example, Arti will presumably look to present or former employees to develop a template for a good worker. But applicants may be very different from current and former workers in a variety of ways. Similarly, the data available for applicants might limit Arti's ability to assess them well on relevant axes, and some groups of applicants may be less well represented than others.[9] In either case, garbage in, garbage out.

But Arti is scarcely unique in this regard since the problem of inaccurate or unrepresentative information plagues human decision-making, and discrimination law is quite forgiving when humans make hiring and firing mistakes because of inaccurate information. For example, Title VII does not hold an employer liable for discrimination under the disparate treatment theory when a human relies on inaccurate information to make

8. Barocas & Selbst, *supra* note 6, at 682 (recounting how a system for sorting medical school applicants carried forward prior discrimination in admissions).

9. To the extent Arti is looking at current or past workers to model good employees and those workers are not diverse, Arti might perpetuate current hiring practices rather than pick better workers in some abstract sense. *See* Bodie, *supra* note 3, at 1013 ("It should not be surprising that trying to predict qualities of good future workers based on the qualities of current workers and existing work culture will not lead to change. In other words, people analytics runs the risk of homosocial reproduction, or replacement of workers with workers that look like them, on a grand scale."); *see also* Alan G. King & Marko J. Mrkonich, "*Big Data" and the Risk of Employment Discrimination*, 68 OKLA. L. REV. 555, 574 (2016) [hereinafter King & Mrkonich] ("For example, if incumbents are older than applicants, then the social-media profile of this older group may differ markedly from that of younger job applicants. Accordingly, an algorithm highly accurate in sorting incumbents for their proficiency may yield applicants notable only for their 'retro' tastes and lifestyles."); Kim, *supra* note 1, at 871–72 ("algorithms will not counteract structural forms of workplace bias" in terms of the way the workplace is currently organized; for example, some kinds of work schedules are particularly difficult for women with childcare responsibilities).

a decision adversely affecting a protected class member. This is the essence of the "honest belief" rule.[10]

Second, it's possible that many traits of good employees are not capable of being captured by the information Arti has at its disposal—no matter how voluminous it is. Further, the more abstract the trait (loyalty, creativity, dedication), the less well are Arti's results likely to map onto some idealized perfect worker. In other contexts, hard data can dwarf "soft variables," even if the latter are in some sense more important,[11] and that may well be true of Arti.

However, that's not a fatal, maybe not even a serious, objection to an employer using Arti for this task. After all, the question is not whether Arti is perfect but rather whether it is as good as or better than current practices. And the antidiscrimination statutes don't really care whether any particular selection device actually improves productivity so long as it does not discriminate.[12] Ultimately, it's not as if we humans have very effective ways of ascertaining abstract traits, and the premise of people analytics is that reliance on data will yield better results than traditional methods of decision-making. This is particularly true of the hiring process, where decisions have usually been made on the basis of very limited (and often imperfect) information. If Arti can identify workers who are likely to be "good" on one or a few axes, that may be sufficient to justify its use as compared to traditional selection processes. And the employer will have post-hiring opportunities to assess more holistic aspects of employee performance.

---

10. *See* Forrester v. Rauland-Borg Corp., 453 F.3d 416, 419 (7th Cir. 2006) (evidence that the defendant's proffered reason was not factually correct was not necessarily probative of pretext because "[a] pretext, to repeat, is a deliberate falsehood. An honest mistake, however dumb, is not . . . ." (citation omitted)); *see also* Johnson v. AT&T Corp., 422 F.3d 756, 762 (8th Cir. 2005) ("[T]he proper inquiry is not whether AT&T was factually correct in determining that Johnson had made the bomb threats. Rather, the proper inquiry is whether AT&T honestly believed that Johnson had made the bomb threats."). But note that perhaps the more egregious the "mistake" the more likely the trier of fact is to find the supposed reason is a pretext for discrimination. *See generally* Ernest F. Lidge III, *Disparate Treatment Employment Discrimination and an Employer's Good Faith: Honest Mistakes, Benign Motives, and Other Sincerely Held Beliefs*, 36 OKLA. CITY U. L. REV. 45 (2011).

11. Laurence H. Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329, 1361 (1971); *see also* Laurence H. Tribe, *Seven Deadly Sins of Straining the Constitution Through a Pseudo-Scientific Sieve*, 36 HASTINGS L.J. 155, 161 (1984) (noting the "pernicious tendency" of cost-benefit analysis to "dwarf soft variables" in constitutional law).

12. The point is simply that the antidiscrimination laws do not require shareholder value maximization; that's a goal that must be reconciled with various legal requirements, including antidiscrimination laws, which may sometimes tend to reduce profits. The statutes do accommodate productivity concerns by allowing neutral practices with a disparate impact to be justified by business necessity. See text beginning *infra* note 89.

In any event, there are reasons to expect Arti to do better than humans by reducing irrationality[13] in employee selection, and that includes performing better than we humans have done in avoiding discrimination.[14] A wide range of research suggests that human bias has considerable negative consequences. This research ranges from audit studies or field experiments showing disparate treatment in real world settings[15] to statistical analyses finding that minorities and women get the

13. Even "rational" selection criteria may be problematic from a performance perspective, and, to the extent that, for example, traditional predictors of success favor whites, they may result in unjustified unfavorable outcomes for minorities. Thus, Google discovered that selection based on such "rational" factors as college attended and grade point average, were not as predictive of success on the job as other factors—such as cognitive abilities, intellectual humility, and the ability to learn—that could be assessed by an appropriate process. Jennifer Alsever, *How AI Is Changing Your Job Hunt,* FORTUNE, (May 19, 2017), http://fortune.com/2017/05/19/ai-changing-jobs-hiring-recruiting/ [https://perma.cc/4WZ2-SABX]; *see also* Josh Constine, *Pymetrics Attacks Discrimination in Hiring with AI and Recruiting Games*, TECHCRUNCH.COM (Sept. 20, 2017), https://techcrunch.com/2017/09/20/perunbiased-hiring/ [https://perma.cc/GE3T-3KG6].

14. *See* Kim, *supra* note 1, at 871 (recognizing that data-driven selection may avoid the problems in the "considerable amounts of bias" inherent in "subjective assessments, intuition, and limited human cognition," but arguing that, since "algorithms are not always neutral either . . . the real question is how the biases they may introduce compare with the human biases they avoid."); Bodie, *supra* note 3, at 1010–11 ("Decisions made by well-meaning people are often flawed by implicit biases that systematically disadvantage historically disadvantaged groups. The ability to analyze accurately what employee traits and skills a business needs to thrive is immensely valuable. And the ability to do that in a way that considers a person's skills accurately without revealing aspects of a person's identity that could trigger bias, whether explicit or implicit, is even more valuable, not just to the business but to the equality project and society more broadly."); *see also* Robert Bolton, *Artificial Intelligence: Could an Algorithm Rid Us of Unconscious Bias?*, PERSONNEL TODAY (Nov. 16, 2017) https://www.personneltoday.com/hr/artificial-intelligence-algorithm-rid-us-unconscious-bias/ [https://perma.cc/4J3E-4J4Z].

15. In these studies, researchers try to directly test the operation of bias by having matched pairs of real or simulated applicants—each pair as similar as possible except for the variable of interest (race or sex)—apply for real-world positions. If one group is more successful than the other, there is reason both to believe that bias exists and that it affects actual decision-making. A significant study in the employment context sent otherwise identical resumes to employers; those using names that did not "sound" African American received more favorable treatment. 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, 991–92 (2004); *see also* David Neumark, Roy J. Bank & Kyle D. Van Nort, *Sex Discrimination in Restaurant Hiring: An Audit Study*, 111 Q.J. ECON. 915, 917-18 (1996). An earlier instance was a study by the Urban Institute that sent matched pairs of black and white testers into the job market, with African Americans faring substantially worse. MARGERY A. TURNER, MICHAEL FIX & RAYMOND J. STRUYK, OPPORTUNITIES DENIED, OPPORTUNITIES DIMINISHED: RACIAL DISCRIMINATION IN HIRING 37–66 (1991). *But see* RICHARD A. EPSTEIN, FORBIDDEN GROUNDS: THE CASE AGAINST EMPLOYMENT DISCRIMINATION LAWS 55–58 (1992). In the legal context, a recent study showed that law firm partners rated work product as worse when the same submission was thought to be written by an African American attorney rather than a Caucasian one. *See* Arin N. Reeves, *Written in Black & White: Exploring Confirmation Bias in Racialized Perceptions of Writing Skills*, NEXTIONS (2014),

short end of the stick across many contexts.[16] To those showings can be added tests of implicit bias that suggest that discriminatory "attitudes" are pervasive.[17] While there are serious questions about the extent to which such tendencies meaningfully influence actual decision-making,[18] the risk

http://nextions.com/wp-content/uploads/2017/05/written-in-black-and-white-yellow-paper-series.pdf [https://perma.cc/AVC6-3NMF].

16. Statistical analyses use retrospective data to seek to hold constant a large number of variables in order to determine whether racial or other bias exists. A dramatic example (albeit not in the employment context) is research showing that the NBA referees were more likely to call fouls on players of a different race than themselves. *See* Joseph Price & Justin Wolfers, *Racial Discrimination Among NBA Referees*, 125 Q.J. ECON. 1859, 1859–60 (2010) (finding statistically significant evidence of own-race bias among NBA referees). More attuned to the employment setting, another study found that store managers were more likely to hire members of their own race than members of another race. Laura Giuliano, David Levine & Jonathan Leonard, *Manager Race and the Race of New Hires*, 27 J. LAB. ECON. 589 (2009).

17. Although there are a host of studies bearing on cognitive bias, the social science research that has perhaps received the most attention is the Implicit Association Test, which purports to measure attitudes at variance with the subjects' expressed views. Hosted at Harvard and available on the Internet, Project Implicit is open to anyone with an Internet connection. PROJECT IMPLICIT, https://implicit.harvard.edu/implicit [https://perma.cc/5GVT-HSKW] (2011). It measures biases (or "implicit attitudes") by comparing how quickly a test taker equates positive and negative words with images of members of different races (and other categories of interest). These results are then compared with the subject's self-reported views on race. The IAT has generated a substantial social science literature analyzing the results of literally hundreds of thousands of visits.

The IAT has been the subject of harsh criticism in the legal academy. The critics have argued that measuring attitudes by millisecond responses to stimuli is inherently flawed and that, even if the test does in some sense identify such attitudes, there is little evidence that they in fact affect real-world decision-making. *See* Amy L. Wax, *Supply Side or Discrimination? Assessing the Role of Unconscious Bias*, 83 TEMP. L. REV. 877, 883–902 (2011); Gregory Mitchell, *Second Thoughts*, 40 MCGEORGE L. REV. 687, 687 (2009); Gregory Mitchell & Philip E. Tetlock, *Facts Do Matter: A Reply to Bagenstos*, 37 HOFSTRA L. REV. 737, 738 (2009); Amy L. Wax, *The Discriminating Mind: Define It, Prove It*, 40 CONN. L. REV. 979, 984–85 (2008); Gregory Mitchell & Philip E. Tetlock, *Antidiscrimination Law and the Perils of Mindreading*, 67 OHIO ST. L.J. 1023, 1023 (2006). But the IAT has also garnered substantial support. *See* Samuel R. Bagenstos, *Implicit Bias, "Science," and Antidiscrimination Law*, 1 HARV. L. & POL'Y REV. 477, 482 (2007); *see also* Jerry Kang & Mahzarin R. Banaji, *Fair Measures: A Behavioral Realist Revision of "Affirmative Action"*, 94 CAL. L. REV. 1063 (2006); Linda Hamilton Krieger & Susan T. Fiske, *Behavioral Realism in Employment Discrimination Law: Implicit Bias and Disparate Treatment*, 94 CAL. L. REV. 997 (2006); Jerry Kang, *Trojan Horses of Race*, 118 HARV. L. REV. 1489 (2005).

18. Recent scholars have both questioned the extent to which implicit bias, at least as measured by the IAT, influences real world decisions. They have also questioned whether the very concept has inappropriately shifted attention from more conscious forms of discrimination. *See* Michael Selmi, *The Paradox of Implicit Bias and a Plea for a New Narrative*, 50 ARIZ. ST. L. REV. 190, 194 (2018) ("[L]abeling nearly all contemporary discrimination as implicit and unconscious is likely to place that behavior beyond legal reach. And it turns out that most of what is defined as implicit bias could just as easily be defined as explicit or conscious bias."); *see also* Samuel R. Bagenstos, *Implicit Bias's Failure*, BERKELEY J. EMPL. & LABOR L. (forthcoming 2018) (manuscript at 9) (available at https://ssrn.com/abstract=3015031) [https://perma.cc/8BTQ-QVZK] ("And the repeated invocation of the con-

of implicit bias and the other studies documenting the extent of discriminatory outcomes might well make Arti seem the safer choice than current reliance on humans.[19]

So we can start out with some optimism about using Arti to select good workers. Implementing its task, Arti could be expected to devise a mechanism that incorporates many factors, familiarly known as an algorithm. That term connotes a multifactored, sophisticated mathematical operation, and that might well be the end result of Arti's operations. But Arti could also devise a far simpler selection device, and its choice between a single criterion and a more complicated algorithm would presumably depend on the data available and Arti's ability to correlate that data with some (or many) measures of productivity.

We will not, of course, know what criteria Arti will choose until we let it operate, but let's explore a few possibilities and examine how the law might react to its actions. Indeed, were this real life, and the employer's General Counsel got wind of Arti's assignment, she might be well advised in light of recent software-related scandals to explore the legal risks entailed.[20] The results of such an analysis may be more than a little surprising.

---

cept of implicit bias by political progressives suggests that old-fashioned intentional discrimination is a thing of the past, when in fact it may simply be better hidden. Indeed, at a moment in history when overt racism—seen in the reaction among some to the election of a black president, and in a significant part of the movement that elected Donald Trump—once again seems a major factor in our public life, the suggestion that implicit bias is the central problem may be particularly misleading."); Ralph Richard Banks & Richard Thompson Ford, *(How) Does Unconscious Bias Matter?: Law, Politics, and Racial Inequality*, 58 EMORY L.J. 1053, 1058 (2009) (arguing that the disparity between IAT results and expressed egalitarian attitudes is equally consistent with unconscious bias and conscious but concealed bias).

19. This might be reinforced by programming Arti to entirely exclude certain factors (prohibited characteristics under the antidiscrimination laws) so that the new-and-improved Arti is literally color- and gender-blind. However, such programming would not necessarily be effective. *See* text beginning *infra* note 77.

Technological efforts short of Arti have been suggested as means of reducing bias. *See* Nancy Leong, *The Race-Neutral Workplace of the Future*, 51 U. CAL. DAVIS L. REV. 719 (2017) (describing various efforts to mask gender during the interview process). *But see* Katharine Zaleskijan, *Job Interviews Without Gender*, N.Y. TIMES, Jan. 6, 2018, https://www.nytimes.com/2018/01/06/opinion/sunday/job-interview-without-gender.html?action=click&pgtype=Homepage&clickSource=story-heading&m [https://perma.cc/HM6M-2Z6B] (arguing that masking for interviews does not address the real problem of workplaces that are not hospitable to women). *See also* EDUARDO BONILLA-SILVA, RACISM WITHOUT RACISTS: COLOR-BLIND RACISM AND THE PERSISTENCE OF RACIAL INEQUALITY IN THE UNITED STATES (4th ed. 2013) (questioning the current "color blind" paradigm).

20. The Volkswagen diesel emissions scandal is the poster child for firms having some controls over software processes. Cary Coglianese & Jennifer Nash, *The Law of the Test: Performance-Based Regulation and Diesel Emissions Control*, 34 YALE J. REG. 33 (2017).

II. ARTI GOING ROGUE

First, consider the most extreme possibility that (shades of Hal!) Arti might go rogue. In other words, suppose Arti decides that a prohibited trait under the antidiscrimination laws is a valid predictor of being a good employee. Despite our societal commitment to equality, this is not far-fetched. Scholars have long recognized "statistical discrimination,"[21] the possibility that employers discriminate not out of any animus but because the protected group in question is (or is perceived to be) less productive (or more expensive) than other workers. For example, an employer may be acting rationally (if illegally) in not hiring disabled workers,[22] older workers,[23] women of child-bearing age,[24] or those disposed to genetic diseases[25] because of perceived greater costs due to potential accommodations, health insurance premiums or shorter job tenure. The claim isn't that such decisions are necessarily correct: the perception that certain workers are more expensive and/or have less job attachment may not be

21. *See, e.g.*, Steven L. Willborn, *The Disparate Impact Model of Discrimination: Theory and Limits*, 34 AM. U. L. REV. 799, 818–19 (1985) ("The statistical theory of discrimination is based on an assumption of market imperfection. This model assumes that employers lack sufficient information to evaluate at a reasonably low cost the productivity potentials of workers. Employers, therefore, substitute readily available proxies such as race, sex, education, or experience for precise, but costly, productivity information. . . . Under the statistical model, employer discrimination reflects not tastes . . . but rather perceptions of reality. That is, racial discrimination reflects employer expectations of the comparative productivity of black and white workers. Under certain conditions, the statistical model of discrimination may explain the persistence of discrimination even in an otherwise perfectly competitive environment."); David A. Strauss, *The Law and Economics of Racial Discrimination in Employment: The Case for Numerical Standards*, 79 GEO. L.J. 1619, 1622–23 (1991); *see also* Stewart J. Schwab, *Is Statistical Discrimination Efficient?*, 76 AM. ECON. REV. 228 (1986); *cf.* Aziz Z. Huq, *Judging Discriminatory Intent*, 103 CORNELL L. REV. (forthcoming 2018) (available at https://ssrn.com/abstract=3033169) [https://perma.cc/8UUS-EZY4] (exploring philosophical and legal implications of statistical discrimination).

22. *See* Sharona Hoffman, *Big Data and the Americans with Disabilities Act*, 68 HASTINGS L.J. 777, 793 (2017); *see also* King & Mrkonich, *supra* note 9, at 581–83.

23. *See* Berhanu Alemayehu & Kenneth Warner, *The Lifetime Distribution of Health Care Costs*, 39 HEALTH SERV. RES. 627 (2004) (health care costs increase with age). In an attempt to deal with one aspect of this risk, the Age Discrimination in Employment Act permits employers to provide lower benefits for older workers so long as they incur equal costs for such benefits. Thus, where a "bona fide employee benefit plan" is involved, an employer will not violate the statute if it either (1) provides its workers equal benefits (in which case there is no discrimination) or (2) provides age-differentiated benefits but incurs equal costs in doing so. 29 U.S.C. § 623(f)(2) (2018).

24. *See* Shannon Weeks McCormack, *Postpartum Taxation and the Squeezed Out Mom*, 105 GEO. L.J. 1323, 1333 (2017) ("[T]he work patterns of the 90% of women who are mothers will frequently diverge from those of men and childless women, leading one expert to observe that '[o]ur economy is divided into mothers and others.'").

25. *See* Ifeoma Ajunwa, *Genetic Data and Civil Rights*, 51 HARV. C.R.-C.L.L. REV. 75 (2016).

true, and, even if it is, such workers may have countervailing advantages.[26] Nevertheless, suppose Arti concluded the contrary when working with a particular set of data, and excluded women of childbearing age.

If a human decision maker drew that gender line, we would label it disparate treatment discrimination and strike it down absent some statutory defense. Indeed, the Supreme Court did exactly that in *Phillips v. Martin Marietta Corp.* with respect to an employer's policy excluding women with pre-school age children.[27] Absent a statutory defense,[28] it has not mattered whether the group in question is in fact different in job related ways.[29] Rather, the individual focus of Title VII bars employers from treating applicants and employees as if they share the traits of the protected class to which they belong. For this reason, antidiscrimination law forbids reliance on prohibited grounds, even if that reliance is statistically rational.[30]

But the policy in *Martin Marietta* was promulgated by humans. Suppose, instead, it is Arti that draws such a line. Shockingly it (or, more accurately, the employer who deploys it) does not seem to have violated the law as the Supreme Court has declared it: put simply, Arti has not engaged in either disparate treatment or disparate impact as those terms have been defined by the Court, which has repeatedly described these two theories as if they comprise the entire universe of "discrimination."[31]

---

26. Indeed, Arti might be the solution instead of the problem. As we've seen, one advantage of nonhuman selection processes is filtering out human biases. Arti might "know" that many stereotypes that influence current decision-making are not true and therefore its selection of employees should be more fact-based and less driven by biases. If we let the Artis of the world proceed without regulation, it would be an empirical question whether the cause of equality would, overall, be advanced or retarded.

27. 400 U.S. 542, 544 (1971) (finding the lower court erred in reading § 703(a) "as permitting one hiring policy for women and another for men—each having pre-school-age children"). The Court, however, remanded for consideration of a possible bona fide occupational qualification defense. 42 U.S.C. § 2000e-2 (2018) (Title VII); *see also* 29 U.S.C. § 623(f)(1) (2018) (ADEA). It did not matter to the Court that the policy at issue in *Martin Marietta* did not discriminate against all women, just those with pre-school age children, which has generally been described as a "sex-plus" theory. *See* Noah D. Zatz, *Disparate Impact and the Unity of Equality Law*, 97 B.U. L. REV. 1357, 1371–72 (2017) [hereinafter Zatz].

28. The following discussion ignores the possibility that a particular act of discrimination may be permitted under one of the exceptions to the antidiscrimination statutes, such as the bfoq. 42 U.S.C. § 2000e-2 (Title VII); 29 U.S.C. § 623(f)(1) (ADEA).

29. See discussion in text beginning *infra* note 70.

30. In this regard, however, Arti might be the solution instead of the problem. As we've seen, one advantage of nonhuman selection processes is filtering out human biases. Arti might "know" that many stereotypes that influence current decision-making are not true and therefore its selection of employees should be more fact-based and less driven by biases. If we let the Artis of the world proceed without regulation, it would be an empirical question whether the cause of equality would be advanced or retarded.

31. Int'l Bhd. of Teamsters v. United States, 431 U.S. 324, 335 n.15 (1977). See discussion in text beginning *infra* notes 34 and 55.

A. *Disparate Treatment and the Limits of "Intent"*

Most obviously, Arti isn't human, so it can't "intend" to discriminate,[32] and Supreme Court precedent requires "intent" or "motive" (the terms are often used interchangeably[33]) for what it labels "disparate treat-

---

I put to one side a third theory, nonaccommodation, which is largely confined to the Americans with Disabilities Act, *see* § 12112(b)(5) (defining discrimination to include "not making reasonable accommodations to the known physical or mental limitations of an otherwise qualified individual with a disability . . . unless such covered entity can demonstrate that the accommodation would impose an undue hardship on the operation of the business of such covered entity"), although it has limited resonance in Title VII with respect to religion, EEOC v. Abercrombie & Fitch Stores, Inc., 135 S. Ct. 2028 (2015), and, arguably, pregnancy, Young v. UPS, 135 S. Ct. 1338 (2015); *see also id.* at 1356–57 (Alito, J., concurring) (while under the first clause of the Pregnancy Discrimination Act "all that matters is the employer's actual intent," the second clause "adds a further requirement of equal treatment irrespective of intent.").

Some view harassment as yet a fourth theory of discrimination, Katherine M. Franke, *What's Wrong with Sexual Harassment?*, 49 STAN. L. REV. 691, 691–92 (1997), but that doctrine fits more-or-less well into traditional disparate treatment since it requires the perpetrator to be motivated by the protected trait of the victim, Oncale v. Sundowner Offshore Servs., Inc., 523 U.S. 75, 80 (1998) ("Title VII does not prohibit all verbal or physical harassment in the workplace; it is directed only at 'discrimination . . . because of . . . sex.' We have never held that workplace harassment, even harassment between men and women, is automatically discrimination because of sex merely because the words used have sexual content or connotations.").

32. Concerns have been raised about using Big Data in employee selection in terms of discrimination being intentionally programmed in by those designing the algorithms. Barocas & Selbst, *supra* note 6, at 692–94. While such concerns are legitimate where humans are involved, the conceit of this article is that the artificial intelligence is making all decisions free of human intervention—beyond providing it with the instruction to choose good employees and providing it with all available data that might bear on that task. Human bias is, by hypothesis, excluded from consideration.

33. *See generally* Charles A. Sullivan, *Disparate Impact: Looking Past the* Desert Palace *Mirage*, 47 WM. & MARY L. REV. 911, 914–16 (2005).

In *Staub v. Proctor Hospital*, 562 U.S. 411 (2011), the Court made a conscious effort to distinguish "motive" from "intent," with both required for employer liability: the relevant agent must possess a discriminatory motive and must also have intended the conduct to cause plaintiff to suffer an adverse employment action; further, that agent must have proximately caused the resulting adverse action. It is not clear whether all agents with both motive and intent can trigger liability for the employer. 562 U.S. at 422 n.4 ("We express no view as to whether the employer would be liable if a co-worker, rather than a supervisor, committed a discriminatory act that influenced the ultimate employment decision."). *See generally* Charles A. Sullivan, *Tortifying Employment Discrimination*, 92 B.U. L. REV. 1431 (2012).

However, the Court soon reverted to its former use of the terms as synonymous. *E.g.*, Young v. UPS, 135 S. Ct. 1338, 1356 (2015) ("Claims of discrimination under [the first clause of the Pregnancy Discrimination Act] require proof of discriminatory intent. Thus, as a result of the first clause, an employer engages in unlawful discrimination under §2000e-2(a)(1) if (and only if) the employer's intent is to discriminate because of or on the basis of pregnancy. If an employer treats a pregnant woman unfavorably for any other reason, the employer is not guilty of an unlawful employment practice under §2000e-2(a), as defined by the first clause of the PDA. And under this first clause, it does not matter whether the

ment" cases. In its seminal decision in *Int'l Bhd. of Teamsters v. United States*, the Court bifurcated the universe of Title VII violations into disparate treatment and disparate impact,[34] and it described disparate treatment as "the most easily understood type of discrimination. The employer simply treats some people less favorably than others because of their race, religion, sex, or national origin. *Proof of discriminatory motive is critical*, although it can in most situations be inferred from the mere fact of differences in treatment."[35] Courts have repeatedly quoted the "proof of discriminatory motive is critical" language,[36] and there are literally thousands of cases that speak of "discriminatory intent" as the sine qua non of a disparate treatment violation.[37] Further, the 1991 Civil Rights Amendments reinforce this focus to the extent they added "motivating factor" liability to the statute.[38] Finally, some Supreme Court Justices have recently stressed the necessity of a wrongful motive for a Title VII disparate treatment violation.[39]

Arti doesn't have any "motives,"[40] which seems to mean that its using a prohibited criterion to select good employees can't be said to violate Title VII's disparate treatment prohibition.[41] In one sense that is hardly

---

employer's ground for the unfavorable treatment is reasonable; all that matters is the employer's actual intent.") (citations omitted); Univ. of Tex. Sw. Med. Ctr. v. Nassar, 570 U.S. 338, 358-59 (2013) ("It would be inconsistent with the structure and operation of Title VII to so raise the costs, both financial and reputational, on an employer whose actions were not in fact the result of any discriminatory or retaliatory intent.").

34. *See* discussion in text beginning at note 54.

35. Int'l Bhd. of Teamsters v. United States, 431 U.S. 324, 335 n.15 (1977) (citation omitted) (emphasis added). The Court went on to write: "Undoubtedly disparate treatment was the most obvious evil Congress had in mind when it enacted Title VII." *Id.*

36. *E.g.*, Smith v. City of Jackson, 544 U.S. 228, 250 (2005); Hazen Paper Co. v. Biggins, 507 U.S. 604, 609 (1993).

37. A Lexis Search yielded 6067 hits ("discriminatory intent" w/p "Title VII"), conducted Dec. 4, 2017.

38. 42 U.S.C. § 2000e-5(g)(2)(B). The division between disparate treatment "motivating factor" liability and disparate treatment liability assessed under *McDonnell Douglas Corp. v. Green*, 411 U.S. 792 (1973), is complex but beyond the scope of this article. *See* Charles A. Sullivan, *Disparate Impact: Looking Past the* Desert Palace *Mirage*, 47 WM. & MARY L. REV. 911, 925–38.

39. *See infra* note 50.

40. *See* Bodie, *supra* note 3, at 1025–26 ("[A] necessary implication of Barocas and Selbst's discussion is that disparate treatment may be impossible to prove because machines are not sentient—they cannot have motives. Decisions can be attributed to algorithms developed over time by the analytics process itself rather than by human design. This kind of discrimination sounds more like disparate impact.")

41. The antidiscrimination laws are directed at "employers" (as well as employment agencies and labor organizations), which are typically business entities rather than natural persons. Speaking of the motives of an artificial intelligence might not look so odd when we routinely describe artificial entities as having "intent" (as in the ubiquitous phrase "legislative intent"). But in the antidiscrimination arena, the cases to date have addressed that question by looking to the

surprising: legal rules first formulated more than forty years ago don't map very well onto Arti's actions. Admittedly, there is an obvious workaround: anthropomorphizing Arti, which would impute to the employer both the prohibited motivation and intent to cause an adverse employment action when Arti classifies on prohibited grounds.[42] That would "solve" the precedent problem[43] but only by creating a legal fiction,[44] one designed to adapt traditional doctrine to the brave new world of Arti.[45] However, it would also raise questions of whether "intent" could not more generally be divorced from traditional views of human motivation, and it would generate a real conundrum as to the correct analysis at the obvious next stage. That is, suppose that, to avoid this problem, Arti is programmed not to use protected traits in its operations. While it would then be race- and gender-blind, faithfulness to its mission would seem to require it

---

motives of human beings within the entity, *see* Staub v. Proctor Hosp., 562 U.S. 411 (2011), discussed *supra* note 33.

42. Employers are responsible for the discriminatory acts of their decision makers, and, as *Staub* suggests, even for discriminatory acts by lower-tier supervisors whose bias is intended to and does proximately cause adverse employment actions. They may or may not also be responsible when decisions result from the bias of co-workers. *Id.* at 422 n.4. And they may be directly responsible when they fail to deal appropriately with discrimination by co-workers or even third parties. Burlington Indus., Inc. v. Ellerth, 524 U.S. 742 (1998) (setting out the liability structure for harassment by supervisors and others). In all these cases, however, there is some human being manifesting the requisite discriminatory motive.

43. However, to hold the employer liable for the actions of Arti would seem to require some modification to the law of agency which, at least for some legal purposes, views a computer program not an "agent" to begin with. The Restatement of Agency provides that a computer program is not capable of acting as a principal or an agent as defined by the common law. At present, computer programs are instrumentalities of the persons who use them. If a program malfunctions, even in ways unanticipated by its designer or user, the legal consequences for the person who uses it are no different than the consequences stemming from the malfunction of any other type of instrumentality. That a program may malfunction does not create capacity to act as a principal or an agent. RESTATEMENT (THIRD) OF AGENCY § 1.04 cmt. e. (AM. LAW INST. 2006); *see also id.* illus. 3 (dog acting as instrumentality for its owner may create liability for owner). *But see* Uniform Electronic Transactions Act, UNIF. Law Comm'n 1999, adopted in multiple states, which provides that "electronic agents" may bind those who use them. While this may make sense in, say, electronic contracting, it would exonerate the employer of disparate treatment liability since neither the human nor her instrumentality would have the necessary discriminatory motive, even if the human had been negligent in its programming Arti.

44. There are those, Professor Willborn among them, who view "motive" as developed in the courts as mostly a legal fiction to begin with, that is, as a label placed on a set of facts post *hoc* when we see certain results. That may be a better description of the reality of discrimination law, but is certainly not how the courts view the enterprise.

45. One could describe this as a nondelegable duty not to discriminate, but deploying such a device immediately runs into a conceptual problem. At least when disparate treatment is concerned, there is no discrimination absent a human with the requisite intent. Thus, any employer duty not to discriminate is satisfied.

to look to "neutral" criteria but ones with a high correlation to the now-off-limits prohibited characteristics.[46]

However, there's a more direct and more textual way to reach the same result without treating Arti as human, although that requires abandoning the Court's bifurcated structure and returning to the text of the statute. The governing language of Title VII (and other antidiscrimination laws[47]) clearly proscribes Arti's use of gender as a selection criterion, and does so entirely without the need for either of the two theories articulated by the Court over the years. The core prohibition of Title VII is § 703, whose two subsections have been viewed as the basis, respectively, of disparate treatment and disparate impact liability.[48] But, entirely apart from the judicial gloss of each subsection, the language of the statute would declare Arti's gender-explicit criterion a violation under both prongs. Section 703 declares it an unlawful employment practice

> (1) to fail or refuse to hire or to discharge any individual, or otherwise to discriminate against any individual with respect to his compensation, terms, conditions, or privileges of employment, because of such individual's race, color, religion, sex, or national origin; or
>
> (2) to limit, segregate, or classify his employees or applicants for employment in any way which would deprive or tend to deprive any individual of employment opportunities or otherwise ad-

46. Were a human to resort to this device, we would likely describe it as pretext. But, absent treating Arti as human, current structures seem to require such action to be analyzed as disparate impact. *See* discussion beginning *infra* note 77.

47. *E.g.*, 29 U.S.C. § 623(a) (2018) (ADEA).

48. Subsection (a) is often said to focus on disparate treatment while, because of its "tend to deprive" language, subsection (b) is frequently cited as the basis for disparate impact liability in the statute as it was passed in 1964. However, subsection (b) would clearly include classifications motivated by a prohibited consideration that actually deprived a class member of employment opportunities. *See* International Union, UAW v. Johnson Controls, Inc., 499 U.S. 187, 197 (1991) ("Section 703(a) . . . prohibits sex-based classifications in terms and conditions of employment, in hiring and discharging decisions, and in other employment decisions that adversely affect an employee's status. Respondent's fetal-protection policy explicitly discriminates against women on the basis of their sex. The policy excludes women with childbearing capacity from lead-exposed jobs and so creates a facial classification based on gender."); *see also* Phillips v. Martin Marietta Corp., 400 U.S. 542, 544 (1971) (finding the lower court erred in reading § 703(a) "as permitting one hiring policy for women and another for men—each having preschool-age children"; the opinion quoted both subsections); *Cf.* Sandra F. Sperino, *Justice Kennedy's Big New Idea*, 96 B.U. L. REV. 1789 (2016) (arguing that subsection (b) is applicable to disparate treatment claims but attributing that largely to Justice Kennedy's majority opinion in *Texas Department of Housing & Community Affairs v. Inclusive Communities Project, Inc.*, 135 S. Ct. 2507 (2015). *See generally* Rebecca Hanner White, *De Minimis Discrimination*, 47 EMORY L.J. 1121, 1150 (1998) ("This does not mean that Section 703(a)(2) cannot also support a disparate treatment claim when an employer intentionally limits, classifies, or segregates its employees for a prohibited purpose. It does mean that when Section 703(a)(2) is relied upon as the basis for a claim, adversity must be shown.").

> versely affect his status as an employee, because of such individual's race, color, religion, sex, or national origin.[49]

In our thought experiment, Arti would be discriminating against women by failing or refusing to hire them under subsection (1)[50] and also classifying them in a way that would deprive them of employment opportunities under subsection (2).[51]

In short, a court ready to strike down Arti's facial gender exclusion rule could ground its reasoning on the language of § 703(a) rather than the decisional law which, to date, has, quite naturally, involved human motivations. But such action would cry out for a theoretical justification, with the most obvious being that Title VII embraces a causal view of what

---

49. 42 U.S.C. § 2000e-2(a) (2018).

50. The argument to the contrary can be found in the dissents in *Texas Department of Housing & Community Affairs v. Inclusive Communities Project, Inc.*, 135 S. Ct. 2507 (2015), which seem to read the "because of" language of both subsections to require discriminatory intent to be the causal agent. For example, Justice Thomas wrote:

> Each paragraph in § 2000e-2(a) is limited to actions taken "because of" a protected trait, and "the ordinary meaning of 'because of' is 'by reason of' or 'on account of.'" Section 2000e-2(a) thus applies only when a protected characteristic "was the 'reason' that the employer decided to act." In other words, "to take action against an individual *because of*" a protected trait "plainly requires discriminatory intent."
>
> No one disputes that understanding of § 2000e-2(a)(1). We have repeatedly explained that a plaintiff bringing an action under this provision "must establish 'that the defendant had a discriminatory intent or motive' for taking a job-related action." The only dispute is whether the same language—"because of"—means something different in § 2000e-2(a)(2) than it does in § 2000e-2(a)(1).
>
> The answer to that question *should* be obvious. We ordinarily presume that "identical words used in different parts of the same act are intended to have the same meaning."

Id. at 2526–27 (Thomas, J., dissenting) (citations omitted); *see also id.* at 2535 (Alito, J., dissenting, joined by the Chief Justice and Justice Scalia) ("Like the FHA, many other federal statutes use the phrase 'because of' to signify what that phrase means in ordinary speech. For instance, the federal hate crime statute, 18 U.S.C. § 249, authorizes enhanced sentences for defendants convicted of committing certain crimes 'because of' race, color, religion, or other listed characteristics. Hate crimes require bad intent—indeed, that is the whole point of these laws. All of this confirms that 'because of' in the FHA should be read to mean what it says.") (citations omitted). Of course, the dissenters in *Inclusive Communities* were not faced with Arti, but it is not impossible to imagine them finding the "plain meaning" of "because" in Title VII to exclude nonhuman discrimination, leaving it to Congress to address the gap in coverage thereby created.

51. Kim, *supra* note 1, at 911–12 argues that this subsection would permit attack on what she calls "classification bias," without the need for invoking traditional disparate impact theory. While Congress obviously did not have data mining in mind in 1964, the language of § 703(a)(2) "sweeps broadly enough to reach unanticipated employer practices that exacerbate or entrench inequality on prohibited bases." She offers a new approach that departs from conventional disparate impact theory in a number of ways. *Id.* at 917.

we call disparate treatment rather than a motivational one,[52] although motivation provides one (but only one) causal mechanism.

That, in turn could have enormous implications for more quotidian cases. Most obviously, the debate about whether discrimination requires conscious motivation or whether an actionable decision can flow from unconscious impulses, usually referred to as "implicit bias," would seem to be resolved were a court to find the employer guilty of a violation when Arti employs explicit gender sorting in its hiring decisions. A violation would simply be a matter of causation although proof that a particular adverse employment action was caused by such biases would remain difficult. Another possible implication is revisiting the notion that discrimination on the basis of traits highly correlated with a prohibited characteristic but not coextensive is not actionable.[53] Might a sufficiently high correlation be viewed as causal? Further, those who feel a need to root Title VII's pro-

52. I am by no means the first to see causation as the key to Title VII violations. Efforts to explain how implicit bias could fit within an intent-centric view of disparate treatment necessarily required a turn to causation. *See, e.g.*, Linda Hamilton Krieger, *The Content of Our Categories: A Cognitive Bias Approach to Discrimination and Equal Employment Opportunity*, 47 STAN. L. REV. 1161, 1168 (1995) ("It would be reasonable to interpret [§ 703's] language as simply requiring proof of causation without proof of intent. In other words, a Title VII claimant would need only establish that his or her protected status 'made a difference' or 'played a role' in a challenged employment decision. This is not, however, how section 703 has been construed."). *Accord* Amy L. Wax, *The Discriminating Mind: Define It, Prove It*, 40 CONN. L. REV. 979, 982-83 (2008); Katharine T. Bartlett, *Making Good on Good Intentions: The Critical Role of Motivation in Reducing Implicit Workplace Discrimination*, 95 VA. L. REV. 1893, 1900 (2009); Michael Selmi, *Proving Intentional Discrimination: The Reality of Supreme Court Rhetoric*, 86 GEO. L.J. 279, 294 (1997). More recently, Professor Zatz argued that "status causation" underlies all the theories of discrimination, including non-accommodation, but operates with different causal mechanisms for each. Zatz, *supra* note 27.

53. Trait discrimination has manifested itself in a number of contexts, including national origin (often with the trait being language and race and sex. *See generally* Zachary A. Kramer, *The New Sex Discrimination*, 63 DUKE L.J. 891, 893 (2014) (modern sex discrimination targets men and women who do not conform to workplace norms); Kimberly A. Yuracko, *Trait Discrimination as Race Discrimination: An Argument About Assimilation*, 74 GEO. WASH. L. REV. 365 (2006) (employers should not be allowed to use invalid trait proxies at all and to use valid ones only if there is no disparate impact); Kimberly A. Yuracko, *Trait Discrimination as Sex Discrimination: An Argument Against Neutrality*, 83 TEX. L. REV. 167, 167 (2004) (trait discrimination should be actionable sex discrimination "only when it stems from gender norms and scripts that are themselves incompatible with sex equality in the workplace"); Tristin Green, *Discomfort at Work: Workplace Assimilation Demands, Social Equality, and the Contact Hypothesis*, 86 N.C. L. REV. 379, 386 (2008) ("[E]mployees should be provided space to signal membership in groups protected by Title VII of the Civil Rights Act through employer accommodation of appearance."); D. Wendy Greene, *Title VII: What's Hair (and Other Race-Based Characteristics) Got to Do With It?*, 79 U. COLO. L. REV. 1355, 1393 (2008) (courts should expand the definition of race to include historical and contemporary understandings); Camille Gear Rich, *Performing Racial and Ethnic Identity: Discrimination by Proxy and the Future of Title VII*, 79 N.Y.U. L. REV. 1134, 1239 (2004) (downplaying the risk that courts will recognize employees' race or ethnicity performance only when it "comports with stereotypical negative representations of minority communities").

scriptions in some flaw of the employer might justify the result in terms of the employer's negligence (or even recklessness) in allowing Arti to proceed in its merry way without the kind of programming that would have prevented its using a protected trait as a selection criterion. That despite the general resistance to viewing Title VII as being even in part negligence based.[54]

B. *Disparate Impact to the Rescue?*

Perhaps the question of the legality of Arti's operations under the disparate treatment theory could be avoided entirely: suppose we assume that the requisite human intent is lacking for a violation under that analysis but look to disparate impact to solve the problem of Arti's explicit gender exclusion. Unfortunately, as suggested at the outset, Rogue Arti also fails to meet the requirements of the latter theory, thus falling into the gap created by the Supreme Court's bifurcation of discrimination into disparate treatment and disparate impact. *Teamsters,* having required intent for disparate treatment claims, went on to distinguish disparate impact claims as "involv[ing] employment practices that are *facially neutral* in their treatment of different groups but that in fact fall more harshly on one group than another and cannot be justified by business necessity. Proof of discriminatory motive . . . is not required under a disparate-impact theory."[55]

In short, while Arti's hypothetical rule excluding women of childbearing age would obviously fall more heavily on women than men since women would be the only ones excluded, that rule is scarcely "facially neutral"; rather, it is facially discriminatory precisely because it classifies on a prohibited ground.[56]

The more highly correlated, the more likely a court will view discrimination on the basis of that trait as being the same as discrimination on the protected basis with which it is correlated. But *Hazen Paper Co. v. Biggins*, 507 U.S. 604 (1993), distinguishing between age discrimination and discrimination to avoid pension vesting, indicates that courts draw distinctions even between highly correlated factors. *See also* EEOC v. Catastrophe Mgmt. Solutions, 837 F.3d 1156 (11th Cir. 2016) (Title VII protects against discrimination on the basis of immutable characteristics, not traits culturally associated with race, such as dreadlocks).

54. *See* David Benjamin Oppenheimer, *Negligent Discrimination*, 141 U. PA. L. REV. 899, 936–44 (1993); Stephanie Bornstein, *Reckless Discrimination*, 105 CALIF. L. REV. 1055 (2017); *see also* Patrick S. Shin, *Liability for Unconscious Discrimination? A Thought Experiment in the Theory of Employment Discrimination Law*, 62 HASTINGS L.J. 67, 89–90 (2010) (exploring the normative question of whether the statute should be read to impose liability on those not consciously motivated to discriminate).

55. Int'l Bhd. of Teamsters v. United States, 431 U.S. 324, 335 n.15 (1977) (emphasis added).

56. In discussing this paper with the author, Professor Romberg argued that the Court couldn't have meant to permit facially discriminatory classifications that fell outside of the disparate treatment/disparate impact bifurcation. While he prefers expanding disparate treatment to reach Rogue Arti, he argues that, as a fallback, disparate impact should be read to include any selection criterion with the requisite impact, even if they are facially discriminatory. In this, he can look for support to the codification of disparate impact by the 1991 Civil Rights Act,

This may require some explanation since a gender disqualification buried in computer code might be thought not to be "facial" to begin with. However, "facial" has been used to embrace not just overt policies of discrimination but covert ones so long as they classify workers on prohibited grounds.[57] The fact that an investigator would have to decipher computer code to find the discrimination is no more relevant than the need to decipher more traditional code words to find they express a prohibited preference.[58] In both cases, there is an express (but covert) policy at work to exclude on the prohibited grounds.[59]

There is yet another reason why disparate impact will not resolve the problem of Rogue Arti: as hypothesized, the gender exclusion seems to be justified by business necessity and therefore is not illegal.[60] The disparate impact theory, as traditionally framed, would not seem to invalidate an employment practice that was based on real differences in, say, productivity between the races or genders (although if it was the result of intentional discrimination it would be illegal disparate treatment). One widely accepted formulation of business necessity is found in the Third Circuit's decision in *El v. SEPTA*,[61] which summarized the standard as requiring an employer relying on the defense to "show that a discriminatory hiring policy accurately—but not perfectly—ascertains an applicant's ability to perform successfully the job in question. In addition, Title VII allows the employer to hire the applicant most likely to perform the job successfully over others less likely to do so."[62] Again, the notion of "rational discrimination" suggests that some acts of discrimination will satisfy normal standards of business necessity.

---

which does not explicitly require facial neutrality. 42 U.S.C § 2000e-(k). The argument, however, is just another way of saying that the current bifurcated paradigm does, not account for Arti, which is the point of this article.

57. The most obvious example is *Ledbetter v. Goodyear Tire & Rubber Co.*, 550 U.S. 618, 634, (2007), *abrogated by* the Lilly Ledbetter Fair Pay Act of 2009, *codified, inter alia*, at 42 U.S.C. § 2000e-5(e)(3)(A), which treated as a "facially discriminatory pay structure" one whose discrimination between blacks and whites was revealed only through multiple regression analysis. Bazemore v. Friday, 478 U.S. 385 (1986).

58. There are currently lawsuits pending challenging alleged employer racial and ethnic preferences for certain kinds of workers from staffing firms. Among the allegations are code words said to reflect those preferences. *See* Kelly Heyboer, *No 'Ghetto People': How Temp Agencies Allegedly Hire Based on Race and Gender*, NJ.COM (Sept. 19, 2016), http://www.nj.com/news/index.ssf/2016/09/no_ghetto_people_how_temp_agencies_allegedly_hire.html [https://perma.cc/72NT-U9SN].

59. Arti's work might be less obvious and less easy to establish in litigation, being hidden in a proprietary algorithm; but intentional discrimination is also frequently kept under wraps for obvious reasons.

60. This requires an important caveat: even when an employer establishes the business necessity for a policy, the plaintiff may nevertheless establish liability by showing an "alternative employment practice." § 703(k)(1)(A) & (C); *see* Jones v. City of Boston, 845 F.3d 28 (1st Cir. 2016). *See* discussion *infra* at note 131.

61. 479 F.3d 232 (3d Cir. 2007).

62. *Id.* at 242.

To see this, consider *International Union v. Johnson Controls*,[63] where at issue was an employer's "fetal protection policy," which excluded women deemed capable of child-bearing from positions in which they would be exposed to elevated levels of lead. The en banc Seventh Circuit majority[64] had viewed this as a disparate impact case, in large part because of the greater flexibility allowed by the business necessity defense than would be true were the bfoq to apply.[65] Not surprisingly given that move, it then found the policy justified by business necessity.[66] The Supreme Court was, therefore, faced with the question whether business reasons could validate an express gender exclusion under disparate impact's business necessity defense. It, however, avoided the question by viewing the policy as a species of disparate treatment, finding intent to discriminate even absent animus since "the absence of a malevolent motive does not convert a facially discriminatory policy into a neutral policy with a discriminatory effect."[67] That analysis meant that the only defense possible was the narrower bona fide occupational qualification exception, not the broader business necessity justification, and the Court held the policy not to be a bfoq.[68]

---

63. 499 U.S. 187 (1991).

64. 886 F.2d 871 (7th Cir. 1989). The court had splintered with some judges applying business necessity and others bfoq. The majority, however, upheld the policy as a business necessity although it would also have found the bfoq defense established.

65. Int'l Union, UAW. v. Johnson Controls, Inc., 886 F.2d 871, 886-87 (7th Cir. 1989) ("We are convinced that the components of the business necessity defense the courts of appeals and the EEOC have utilized in fetal protection cases balance the interests of the employer, the employee and the unborn child in a manner consistent with Title VII. The requirement of a substantial health risk to the unborn child effectively distinguishes between the legitimate risk of harm to health and safety which Title VII permits employers to consider and the '[m]yths or purely habitual assumptions' that employers sometimes attempt to impermissibly utilize to support the exclusion of women from employment opportunities. Likewise, the requirement that the risk of harm to offspring be substantially confined to female employees means that a fetal protection policy applying only to women recognizes the basic physical fact of human reproduction, that only women are capable of bearing children. Finally, the employee's option of presenting less discriminatory alternatives to a fetal protection policy assures that these policies are only as restrictive as necessary to prevent the serious risk of harm to the unborn child."). Other courts had also approved the use of the business necessity defense for such policies. *E.g.*, Wright v. Olin Corp., 697 F.2d 1172 (4th Cir. 1982) (remanding for consideration of that defense).

66. The Seventh Circuit's decision was rendered during the brief period after *Wards Cove Packing Co. v. Atonio*, 490 U.S. 642 (1989), and before the effective date of the 1991 Civil Rights Act during which the business necessity defense had been watered down and employees had the burden of persuasion that business necessity was not applicable. *See generally* Charles A. Sullivan, *Disparate Impact: Looking Past the* Desert Palace *Mirage*, 47 WM. & MARY L. REV. 911, 960-64 (2005).

67. 499 U.S. at 188.

68. The *Johnson Controls* holding was codified by the Civil Rights Act of 1991, which added § 703(k)(2) to Title VII: "A demonstration that an employment practice is required by business necessity may not be used as a defense against a claim of intentional discrimination under this title."

While *Johnson Controls* was clear that the policy at issue was subject to disparate treatment (together with its narrower bfoq defense) not disparate impact analysis (and its broader business necessity defense), it was equally clear that the facial discrimination there reflected the intent necessary for a disparate treatment violation. A similar analysis could be undertaken with respect to *City of Los Angeles Department of Water & Power v. Manhart* the quintessential rational discrimination case. There the Court found disparate treatment in a policy of requiring women to contribute more to their pensions than similarly situated men because women, as a class, lived longer than men, as a class, and therefore required more contributions to ensure equal monthly benefits on retirement.[69]

With no statutory defense available, the policy was held illegal. But suppose the case is viewed from a disparate impact lens. It certainly seems plausible that the employer could establish a business necessity for requiring additional funding for employees who impose additional costs.

The more-than-a-little surprising conclusion is that, under normal analysis, neither the disparate treatment nor impact theories seem to bar Arti's explicit gender exclusion. Since it's clear that we shouldn't permit such practices, there's something wrong with an analytic framework that leads to such counterintuitive results. As we saw with disparate treatment, there is a way to avoid this conclusion, but, again, the alternative would reject the Court's bifurcation of discrimination into disparate treatment and disparate impact. This time, perhaps a more obvious theory is on offer, one that draws from *Phillips v. Martin Marietta* as well as both *Johnson Controls* and *Manhart*. That theory would make facial discrimination impermissible regardless of motive and regardless of the requirements of the impact.[70]

However, a ban on "facial discrimination" is entirely consistent with conventional disparate treatment theory since in all three cases (maybe in

69. 435 U.S. 702 (1978). The employer required larger contributions for pension funding from women than for men because the longer life expectancy of females meant greater actuarially-projected benefit payments. The Court treated this difference as factually true:

> This case does not, however, involve a fictional difference between men and women. It involves a generalization that the parties accept as unquestionably true: Women, as a class, do live longer than men. The Department treated its women employees differently from its men employees because the two classes are in fact different. *Id.* at 707–08.

Nevertheless, the policy discriminated against women in violation of Title VII because "all individuals in the respective classes do not share the characteristic that differentiates the average class representatives," *id.* at 708, and Title VII requires an individual focus. *Manhart* is conventionally viewed as a variety of disparate treatment discrimination, *See* CHARLES A. SULLIVAN, CASES & MATERIALS ON EMPLOYMENT DISCRIMINATION (2017) (leading off Chapter 2, Systemic Disparate Treatment), where the requisite intent is obvious from the classification, albeit animus did not underlie the policy.

70. This seems to be what Barocas & Selbst mean when they write "Disparate treatment recognizes liability for both explicit formal classification and intentional discrimination." Barocas & Selbst, *supra* note 6, at 694.

all cases where the explicit classification is not drawn by a nonhuman such as Arti[71]), the prohibited intent results in the facial distinction. And that seems to be the correct reading of the cases. While neither *Martin Marietta* nor *Manhart* casts much light on the question since neither mentions either motive or intent, *Johnson Controls* clearly viewed facial discrimination as simply a variety of disparate treatment. At one point, it wrote, "The bias in *Johnson Controls*' policy is obvious,"[72] and, in a more extended passage was even clearer:

> [T]he absence of a malevolent motive does not convert a facially discriminatory policy into a neutral policy with a discriminatory effect. Whether an employment practice involves disparate treatment through explicit facial discrimination does not depend on why the employer discriminates but rather on the explicit terms of the discrimination. In *Martin Marietta,* . . . the motives underlying the employers' express exclusion of women did not alter the *intentionally discriminatory character* of the policy. Nor did the arguably benign motives lead to consideration of a business necessity defense. The beneficence of an employer's purpose does not undermine the conclusion that an explicit gender-based policy is sex discrimination under § 703(a) and thus may be defended only as a bfoq.[73]

The thrust of this passage is that intent to discriminate (i.e., to draw a gender line) is critical but that that animus is not necessary: it is enough that the actor's intended to distinguish between workers on a prohibited ground regardless of whether the impetus was benign or malignant.[74]

---

71. Until the advent of Arti, the possibility of facial discrimination that was not wrongly motivated was limited to somewhat bizarre hypotheticals. For example, suppose an employer believed a certain organization's members caused problems in the workplace, perhaps because of beliefs regarding white supremacy, and banned members of that organization without knowing that it was a religion. *See* Peterson v. Wilmur Commc'ns, Inc., 205 F. Supp. 2d 1014, 1015 (E.D. Wis. 2002).

72. Int'l Union, UAW v. Johnson Controls, 499 U.S. 187, 197 (1991).

73. 499 U.S. at 199-200 (emphasis added). Whether this is consistent with equal protection principles is another question. *See* Wayte v. United States, 470 U.S. 598, 608 n.10 (1985) ("A showing of discriminatory intent is not necessary when the equal protection claim is based on an overtly discriminatory classification.") (citing Strauder v. West Virginia, 100 U.S. 303 (1880)).

74. Given the absence of any discriminatory motive, there could be no claim of systemic disparate treatment, which, like its individual disparate treatment cousin, requires intent to discriminate although in systemic cases the intent is typically inferred from statistical outcomes that seem unlikely to have occurred unless selection processes were driven by prohibited motives. *See, e.g.*, Hazelwood Sch. Dist. v. United States, 433 U.S. 299 (1977); Int'l Bhd. of Teamsters v. United States. 431 U.S. 324 (1977). *See generally* Noah D. Zatz, *Introduction: Working Group on the Future of Systemic Disparate Treatment Law*, 32 BERKELEY J. EMP. & LAB. L. 387 (2011); Tristin K. Green, *The Future of Systemic Disparate Treatment Law*, 32 BERKELEY J. EMP. & LAB. L. 395 (2011); Melissa Hart, *Civil Rights and Systemic Wrongs*, 32 BERKELEY J. EMP. & LAB. L. 455 (2011); Michael Selmi, *Theorizing Systemic Disparate Treatment*

In short, to invalidate Rogue Arti's gender exclusion by treating it as a facial classification achieves the goal of not permitting artificial intelligence to adopt employment practices that would be illegal if deployed by humans, but it does require some tinkering with precedent. If we're willing to do that, the simplest approach would be to replace the current treatment/impact bifurcation but not with a facial discrimination epicycle; rather, as suggested, the Court could simply adopt a pure causation structure. Indeed, given the statutory language, situating a "facial-discrimination" rule in Title VII's text would ultimately have to be in the "because of" language and given that, there's no obvious reason to extend "because of" beyond intent but only to facial discrimination and no further.

To this point, then, the normative takeaway from our thought experiment is that current views of disparate treatment liability reflect a cramped version of what Title VII is most naturally read to prohibit. Motive/intent are one, but only one, causal path that employers can tread.[75]

III. BEYOND ROGUE ARTI

Beyond the borders of our thought experiment, Rogue Arti is unlikely to pose a serious problem. Employers, acting to limit risk or merely to do the right thing, are not likely to let Arti go about its work with no safeguards. Thus, while it may be literally impossible (given Arti's mission and the data it has access to) to prevent it from "knowing" the race, sex, age, and even disability status of the individuals with whom it deals,[76] it can be programmed to avoid directly using those criteria in its operations.

Law: After Wal-Mart v. Dukes, 32 BERKELEY J. EMP. & LAB. L. 477 (2011). As applied to Arti's actions, a systemic disparate treatment case might appear from a comparison of the employer's inputs and output, but any inference of underlying discriminatory intent would be rebutted by the employer's pointing to Arti as the cause.

75. Causation is also operative in the disparate impact context, although the law focuses only on the employer's contribution to what is a causal setting that leads to the disproportionate underrepresentation of minorities and women in the workforce. Ramona Paetzold & Steven Willborn, *Deconstructing Disparate Impact: A View of the Model Through New Lenses*, 74 N.C. L. REV. 325, 353–355 (1996) ("Ordinary disparate impact cases, then, view causation with blinders. The law treats the employer's criterion as the cause of a disparity, even though it may be only one of a wide array of factors necessary to produce the disparity. Ordinary disparate impact cases view causation with blinders, not because the cases arise in a single-cause context, but because they ignore causes external to the employer that contribute to the impact. The blinders necessarily mean that employers may be held legally responsible for impacts that are "'caused' in substantial part by factors external to the employers."). *See also* Zatz, *supra* note 27, at 1358 ("A disparate impact claim's statistical comparison of group outcomes provides evidence that individuals have suffered status causation. Group outcomes are constructed by aggregating individual outcomes. Disparities between group outcomes can emerge only if many individual group members suffer harm because of their protected status (status causation). But not all group members suffer this injury; it is spread unevenly within the group. The statistical evidence demonstrates that some individuals suffered discrimination's injury, but it does not identify which individuals.").

76. I use "knowing" in the sense that the data is available to Arti even if it is instructed not to use it. That essentially maps onto the current world where em-

As has been suggested, that decision would essentially shift the focus of the antidiscrimination laws from disparate treatment, with its single-minded concern for motive, to disparate impact, in which biased motives are irrelevant. Suppose Arti plays it straight (either left to its own devices or as a result of prophylactic programming prohibiting reliance on protected characteristics) and does not use such characteristics as part of its selection process. In still pursuing good employees, the most likely scenario is that Arti will use proxies for the forbidden traits (second-best criteria) to achieve results that approximate what it would have done had not sex been ruled out-of-bounds.[77] If a human were to undertake this exercise, we might well talk of "masking" her true motive,[78] but we've seen that Arti has no motive; it seeks only to find the best workers within the limits of its data and the constraints of its programming. Thus, if it uses a criterion to identify those likely to be good workers, it is simply following the data.

To see this, imagine that, after looking at all the data, Arti concludes that the only criterion that it could measure effectively in terms of performance is job tenure. Given the costs of employee turnover,[79] that might be a valuable contribution even if criteria more closely connected to actual job performance would theoretically be preferable. After all, objective measures on an individual basis may not exist in the data set, and Arti may determine that supervisor evaluations are not reliable and/or aren't predicted by the data available at hiring.

---

ployers usually know protected characteristics even though they are not generally supposed to take them into account.

77. If Arti would perceive gender as a useful sorting criterion, but is prohibited from using gender per se, it will turn to "gender neutral" factors that approximate the result that it would have reached when gender per se was on the table. Like the Target example, it might find purchasing decisions that correlate with pregnancy. Or it might exclude people who took extended leaves from work, a criterion that would presumably exclude many women of childbearing age. As Professor Kim notes in *Auditing Algorithms for Discrimination*, 166 U. PA. L. REV. ONLINE 166, 193 (2017) [hereinafter *Auditing Algorithms*], "[i]n any sufficiently rich dataset, proxy variables likely exist that closely correlate with [protected] characteristics, permitting implicit sorting on these bases."

It is for this reason that some commentators view calls for transparency algorithms to be an inadequate solution to problems posed by such use. *See* Joshua A. Kroll et al., *Accountable Algorithms*, 165 U. PA. L. REV, 633, 633–34 (2016) (challenging the "dominant position" that transparency will solve" such problems both because it is often impracticable and sometimes undesirable and technological innovations allow the design of "decisionmaking algorithms that . . . better align with legal and policy objectives"); *cf.* Kim, *Auditing Algorithms, supra* note 77, at 197 (arguing in favor of retaining audits of outcomes to detect discrimination).

78. Bodie, *supra* note 3, at 1014. Use as an example of masking bias the use of zip codes to screen out minority candidates.

79. *See* Joni Hersch & Jennifer Bennett Shinall, *Something to Talk About: Information Exchange under Employment Law*, 165 U. PA. L. REV. 49, 89 n.137 (2016) (reporting studies reflecting that replacement costs—direct and indirect—can exceed the cost of an employee's annual salary, sometimes by multiples of it).

In any event, suppose job tenure is Arti's criterion for good workers.[80] Presumably, Arti would then[81] canvass its data to determine what factor or factors predict[82] longer length of services. For example, Arti might discover that workers who live closer to the workplace were likely to stay longer. That would mean that "good" is highly correlated with place of residence at hiring relative to the workplace. Under this arrangement, Arti might then hire applicants residing in nearby zip codes and reject those living further away.[83]

Of course, this is greatly oversimplified: Arti is almost certain to devise multiple criteria for good employees and will similarly develop more complicated ways of ascertaining who is most likely to satisfy these criteria. But for purposes of our thought experiment, let's stick to the simplest model: "good" equals length of employment and length of employment is best predicted by place of residence at hiring.

Correlation between residence and longevity may be caused by many things, some of which may conform to our intuitive notions of good workers while others may not. For example, workers may stick around because of the short commute rather than any intrinsic loyalty to the company or love of the work. Similarly, those with longer commutes may tend to leave more quickly precisely for that reason. If Arti is doing its job right, however, it has identified the only criterion for which the data can predict job success (as indicated by length of service) and implemented a hiring system to maximize that success.[84]

In this example, Arti's selections make at least some sense in human terms; that is, we can see a causal connection between the criterion and the performance measure selected. But note that the ability to discover associations with no obvious causal connection is usually viewed as a feature,[85] not a bug,[86] of people analytics, and Arti might well have come up

80. Some might call this a "loyalty" measure, which might make it appear that Arti is looking for more abstract job traits, but that's more a label than a criterion since there are obvious problems with measuring "loyalty" solely by length of service.

81. "Then" suggests a temporal hierarchy of operations that probably doesn't exist since Arti is recursively looking over the data to see what criteria it can assess while simultaneously specifying the metrics for predicting when a candidate is likely to be successful.

82. It's probably more accurate to speak of factors that "correlate" with the relevant criteria, but the point of the exercise is to select workers who will be good in the future, so "predict" seems appropriate.

83. There might be a chicken-egg problem here with more committed workers moving closer to their workplace after being employed, but, presumably, Arti will figure that out.

84. Again, Arti may be wrong, and its hiring algorithm may be a failure. But, by hypothesis, Arti has done the best it could with the available data.

85. Bodie, *supra* note 3, at 969–71 ("Analytics is a term often used in a business context to describe the discovery of meaningful patterns in data, also known as knowledge discovery in data. . . . Data mining usually does not begin with a hypothesis, but instead uses a variety of tools to generate hypotheses and test them against the available data. . . . Data analytics are popular within the HR community

with a criterion—say, taste in music—with no intuitive causal connection to the performance measure it correlates with.[87]

Disparate impact analysis under Title VII[88] generally proceeds in three steps.[89] In the first, plaintiff must identify a "particular employment practice"[90] with an adverse racial impact.[91] If plaintiff is successful, defendant then has the opportunity to establish the affirmative defense that the

---

because they are efficient and effective at using information to find or identify groups or individuals who have desirable skills, attributes, needs, or tastes.").

86. The "bug" part may be the time frames for any such correlations. King & Mrkonich, *supra* note 9, at 562, suggest that causal explanations are likely to be more stable across time than mere correlations. Thus, they report a firm's discovery that expert coders also favor a particular Japanese manga site; that correlation might last only as long as the site in question remains "hot." The authors view this as the Achilles heel of big data in employment: "there is no reason the algorithm that bests fits the data on Monday will do so on Tuesday. . . . Because there is no understanding of *why* the correlation exists, there is no basis for surmising how long it will persist." *Id.* at 578 (emphasis in original).

87. *See* James Grimmelmann & Daniel Westreich, *Incomprehensible Discrimination*, 7 CAL. L. REV. ONLINE 164 (2017) [hereinafter Grimmelmann & Westreich].

88. Michael Selmi, in *Was the Disparate Impact Theory a Mistake?*, 53 UCLA L. REV. 701 (2006), argues that disparate impact has never functioned well outside the context in which it was created—the use of paper-and-pencil test results as excluders of applicants. The analysis that follows suggests he may be correct.

89. Age Discrimination in Employment Act cases have a somewhat different analysis because of the statute's "reasonable factors other than age" defense. *See* Meacham v. Knolls Atomic Power Lab., 554 U.S. 84 (2008); 29 C.F.R. § 1625.7(e) (2018).

90. Civil Rights Act of 1991 § 703(k)(1)(A)(i), 42 U.S.C. § 2000e-2(k)(1)(A)(i) ("An unlawful employment practice based on disparate impact is established under this title only if—(i) a complaining party demonstrates that a respondent uses a particular employment practice that causes a disparate impact on the basis of race, color, religion, sex, or national origin and the respondent fails to demonstrate that the challenged practice is job related for the position in question and consistent with business necessity."). Subsection (k)(1)(B) allows the plaintiff to challenge an entire "decision-making process" rather than a particular practice if it "can demonstrate to the court that the elements of a respondent's decision-making process are not capable of separation for analysis." Section 701(m) defines "demonstrates" as carrying the burden of production and persuasion.

The place of residence example assumes that Arti is using a single criterion with a disparate impact. In more realistic models, the employer's algorithm would use a combination of factors and questions would arise as to which, if any, of those factors had the requisite impact. For example, Arti might construct a regression equation with good job performance as the dependent variable and a variety of data as the independent variables. Whether a plaintiff would be required to challenge the entire algorithm or could focus on particular aspects of it would undoubtedly become an issue. To the extent that factors relied on aren't separate pass/fail screens but interrelate with each other, the entire algorithm might seem the appropriate unit. *See* Chin v. Port Auth. of N.Y. & N.J., 685 F.3d 135, 154–55 (2d Cir. 2012) (while a promotion process formally involved three steps, the first two steps could not be separated from the rest for statistical analysis since both played an indeterminate role); McClain v. Lufkin Indus., 519 F.3d 264, 279 (5th Cir. 2008) (no error in analyzing several employment practices as one where the seniority system and the treatment of absenteeism were subject to considerable managerial discretion).

practice is justified by business necessity and job relation.[92] Finally, if defendant is successful in carrying its burden, the plaintiff may nevertheless prevail by showing that the employer refused to adopt an "alternative employment practice" that would equivalently meet its needs with lesser adverse impact.[93]

With respect to the first step, if a plaintiff proves that Arti looks to job tenure as a measure for good employees and residence proximity to the workplace was the predictor (the "particular employment practice"), an adverse racial impact on African Americans would often exist given housing segregation patterns, depending on the location of the workplace and the demographics of the surrounding areas.[94]

On the other hand, the data is presumably available to determine if each independent variable has a disparate impact (i.e., the selection process is capable of being separated for analysis), and a plaintiff might well prefer to attack them separately; after all, while they may well predict job performance when their values are combined, each considered separately may not be job-related. However, Arti hasn't used the factors separately but as part of its single selection process, so perhaps the algorithm can't "be separated for analysis" in that sense. And there has always been a tension with regard to what constitutes the relevant unit for impact analysis. Traditionally, each "test" has been so viewed, although each question on the test could be examined. In *Dothard v. Rawlinson*, 433 U.S. 321, 323-26 (1977), the Court considered height and weight requirements as a single selection criterion although they could have been separately analyzed. And to add a final complication, an employer is not required to prove the defense for any component of a selection process that it can show does not have the requisite impact, § 703(k)(1)(B)(ii) (if the employer "demonstrates that a specific employment practice does not cause the disparate impact, the [employer] shall not be required to demonstrate that such practice is required by business necessity."); this creates the possibility that the employer might sometimes want to deconstruct its own algorithm.

In any event, the extent to which an algorithm can be deconstructed in litigation is sure to be a focus of concern, and these determinations may be influenced by trade secret considerations. Although Arti is imagined to be owned by the employer, real world employers are likely to use third-party software, whose owners are likely to resist disclosure of the algorithms used. *See* EEOC v. Kronos Inc., 694 F.3d 351 (3d Cir. 2012). *See generally* King & Mrkonich, *supra* note 9, at 557–58.

91. When an identified practice will be found to have such an impact can also raise issues. *See* Ifeoma Ajunwa, Sorelle Friedler, Carlos Scheidegger & Suresh Venkatasubramanian, *Hiring by Algorithm: Predicting and Preventing Disparate Impact*, SSRN pp. 11–13 [listed as a not-to-be-cited draft] (discussing, inter alia, applying the EEOC's four-fifths rule to data mining).

92. § 703(k)(1)(A).

93. § 703(k)(1)(A) ("An unlawful employment practice based on disparate impact is established under this title only if . . . the complaining party [demonstrates the availability of] . . . an alternative employment practice and the respondent refuses to adopt such alternative employment practice." *See infra* note 131.

94. *See* Newark Branch, NAACP v. Town of Harrison, 940 F.2d 792, 800 (3d Cir. 1991). Professor Kim reports an instance in which a company elected not to use distance from work as a predictor of job tenure as part of its hiring algorithm precisely "because it understood that housing patterns are correlated with race and that relying on that correlation might result in discrimination." Kim, *supra* note 1, at 873.

Although in some cases the correlation may be so high that one might question whether it's really the prohibited trait that is driving the bus,[95] we've seen that with Arti there can be no finding of prohibited motive or intent. That leaves the question of whether Arti's criterion is justified by business necessity. And, we've noted that, in the people analytics context, such an impact seems, at first blush, justifiable under traditional analysis. That point is pursued at more length here.

As Barocas and Selbst write, "the very point of data mining is to provide a rational basis upon which to distinguish between individuals and to reliably confer to the individual the qualities possessed by those who seem statistically similar."[96] In other words, data mining seeks to discover statistically significant correlations between the "target variables" (the sought-after characteristics of better workers) and the traits of potential workers to the extent they can be found in the data. And, assuming that the target variable is itself job-related (say, job tenure in our example), the algorithm will identify traits that correlate with that variable—even though there may be no causal explanation available for why that might be true. Thus, these commentators conclude that traditional validation techniques are likely to be satisfied: "Data mining will likely only be used if it is actually predictive of something, so the business necessity defense solely comes down to whether the trait sought is important enough to job performance to justify its use in any context."[97]

There are two main critiques of this approach, one more conceptual and one more technical but both taking aim at the idea that correlation is sufficient to validate a criterion. The more conceptual article is by Grimmelmann & Westreich, who argue that blind correlations should not be accepted as validating the criteria used.[98] They propose that defendant be

95. For example, the Supreme Court has repeatedly had to deal with the question of whether legislative redistricting decisions are race-motivated (unconstitutional absent a compelling state interest) or politically-motivated (constitutional at the moment). *See* Vieth v. Jubelirer, 541 U.S. 267 (2004) (Scalia, J., plurality opinion). The Supreme Court, although granting certiorari, avoided deciding the question of whether some political gerrymanders on standing grounds. *Gill v. Whitford*, 138 S. Ct. 1916 (2018). In any event, given that African-Americans vote very heavily Democratic there is a very strong correlation between the political and the racial possibilities. *See* Cooper v. Harris, 137 S. Ct. 1455, 1473 (2017) ("Getting to the bottom of a dispute like this one poses special challenges for a trial court. In the more usual case alleging a racial gerrymander—where no one has raised a partisanship defense—the court can make real headway by exploring the challenged district's conformity to traditional districting principles, such as compactness and respect for county lines. . . . But such evidence loses much of its value when the State asserts partisanship as a defense, because a bizarre shape—as of the new District 12—can arise from a "political motivation" as well as a racial one. And crucially, political and racial reasons are capable of yielding similar oddities in a district's boundaries. That is because, of course, 'racial identification is highly correlated with political affiliation.'") (citations omitted).

96. *Supra* note 6, at 677.

97. *Id.* at 709.

98. *See* Grimmelmann & Westreich, *supra* note 87.

required to supplement its statistical proof of validation by providing a causal explanation for the result.[99] That is, the defendant's business necessity burden "requires it to show not just that its model's scores are not just *correlated* with job performance but *explain* it."[100] The authors end with a cri de coeur: "Applicants who are judged and found wanting deserve a better explanation than, 'The computer said so.' Sometimes computers say so for the wrong reasons—and it is employers' duty to ensure that they do not."[101]

The more technical critique is found in King & Mrkonich, who believe that satisfying defendant's burden of establishing job relatedness may be Big Data's greatest challenge. Some of Big Data's most vocal advocates contend Big Data is valuable precisely because it crunches data that are ubiquitous and *not* directly job-related . . . . The employer's reliance on the algorithm may be job-related, but the algorithm itself is measuring and tracking behavior that has no direct relationship to job performance. Its value derives solely from a correlation between this [the identified] behavior and job performance. The legal question is whether an employer can meet its burden of proving job-relatedness with evidence that is strictly correlational.[102]

In the job tenure example, it is certainly arguable that longevity is a poor measure of what makes someone a good employee so that Arti's single-minded (so to speak) focus on predictors of that criterion is problematic and perhaps challengeable on that ground.

Indeed, King & Mrkonich suggest looking to the Uniform Guidelines on Employee Selection Procedures[103] as a basis for assessing a system of people analytics.[104] The Guidelines trace back to the early days of Title

---

99. *Id.* at 170.

100. *Id.*

101. *Id.* at 177. Professor Willborn suggests that this may be misguided since society often uses predictors whose relationship to the target variable is not very direct. He argues that the LSAT for law school admissions is a good example of a test that is a good predictor of first year law school performance, *but see* 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, 208 (2016) (the "LSAT predicts more weakly, and UGPA more powerfully, than commonly assumed"), although it doesn't map directly onto law school studies or examination norms. Grimmelmann et al. might respond that the LSAC, which creates and administers the LSAC, has a well-developed rationale for what it tests, a rationale rooted in skills presumably needed for successful law study. Law School Admission Test (LSAT): About the LSAT, LSAC (last visited Apr. 5, 2018) https://www.lsac.org/jd/lsat/about-the-lsat [https://perma.cc/WDS6-WTYX].

102. King & Mrkonich, *supra* note 9, at 571 (emphasis added); *see also* Kim, *supra* note 1 ("If a statistical correlation were sufficient to satisfy the notion of job-relatedness, the standard would be a tautology rather than a meaningful legal test.").

103. 29 C.F.R. § 1607 (2018).

104. King & Mrkonich, *supra* note 9, at 571.

VII and, although they may not have the force of law,[105] they have repeatedly been viewed as authoritative by courts deciding employment discrimination decisions.[106] While they have been used mainly for the validation of traditional paper-and-pencil tests with a disparate impact,[107] the Guidelines broadly apply to any "selection procedure."[108] That means that they should in theory be applied to Arti's criterion. Indeed, just as I've argued that there is no good reason to treat Arti differently from a human when it uses explicit race or gender classifications, there is no good reason to treat its use of a neutral criterion differently than if human agency devised the same selection criterion.

---

105. The original EEOC Guidelines were viewed as "entitled to great deference" by *Griggs v. Duke Power Co.*, 401 U.S. 424, 433–34 (1971), a point confirmed by *Albemarle Paper Co. v. Moody*, 422 U.S. 405, 430–31 (1975), which noted that the "[g]uidelines draw upon and make reference to professional standards of test validation established by the American Psychological Association." While "not administrative regulations promulgated pursuant to formal procedures established by the Congress . . . they do constitute '[t]he administrative interpretation of the Act by the enforcing agency,' and consequently they are 'entitled to great deference'." (quoting *Griggs*).

The original EEOC Guidelines were revised and replaced in 1978 by the current Uniform Guidelines, representing the collective view of the EEOC and a number of other federal agencies (including the Department of Labor, the Civil Service Commission and Department of Justice), have been treated similarly. *E.g.*, Gulino v. N.Y. State Educ. Dep't, 460 F.3d 361, 383–84 (2d Cir. 2006) ("[T]hirty-five years of using these Guidelines makes them the primary yardstick by which we measure defendants' attempt to validate" a test). *But see* Alfred W. Blumrosen, *Affirmation of Affirmative Action Under the Civil Rights Act of 1991*, 45 RUTGERS L. REV. 903, 910 (1993) (arguing for even greater deference because some of the participating agencies had "substantive rulemaking authority," the Guidelines "were adopted after notice and comment procedures," and *Chevron, U.S.A., Inc. v. National Resources Defense Council, Inc.* now gives more weight to agency statements which are adopted through rulemaking.).

106. The Guidelines have been cited in more than 300 cases, including a number of Supreme Court decisions. Lexis Advance search, conducted Dec. 10, 2017.

107. *See generally* Mark S. Brodin, *Ricci v. DeStefano: The New Haven Firefighters Case & the Triumph of White Privilege*, 20 S. CAL. REV. L. & SOC. JUST. 161, 216 (2011).

108. 29 C.F.R. § 1607.3(A) (2018) ("The use of any selection procedure which has an adverse impact on the hiring, promotion, or other employment or membership opportunities of members of any race, sex, or ethnic group will be considered to be discriminatory and inconsistent with these guidelines, unless the procedure has been validated in accordance with these guidelines. . . ."). "Selection procedure" is in turn defined broadly to include "[a]ny measure, combination of measures, or procedure used as a basis for any employment decision," and includes "the full range of assessment techniques from traditional paper and pencil tests, performance tests, training programs, or probationary periods and physical, educational, and work experience requirements through informal or casual interviews and unscored application forms." 29 C.F.R. § 1607.16(Q) (2018). This suggests that Professor Kim is incorrect in viewing the Guidelines as "simply irrelevant" to people analytics, *supra* note 1, at 999, although she is surely correct that they were not developed with an eye to this kind of selection device.

The Guidelines recognize three kinds of validation: criterion, content, and construct.[109] All three require evidence of a sufficient relationship between the selection device and the job performance it is designed to predict:

> Evidence of the validity of a test or other selection procedure by a criterion-related validity study should consist of empirical data demonstrating that the selection procedure is predictive of or significantly correlated with important elements of job performance. Evidence of the validity of a test or other selection procedure by a content validity study should consist of data showing that the content of the selection procedure is representative of important aspects of performance on the job for which the candidates are to be evaluated. Evidence of the validity of a test or other selection procedure through a construct validity study should consist of data showing that the procedure measures the degree to which candidates have identifiable characteristics which have been determined to be important in successful performance in the job for which the candidates are to be evaluated.[110]

With respect to the content[111] and construct[112] techniques, validation requires an employer to conduct a "job analysis,"[113] as a prelude to constructing a test, an exercise that seems largely irrelevant to Arti's opera-

---

109. *See generally* RAMONA L. PAETZOLD & STEVEN L. WILLBORN, THE STATISTICS OF DISCRIMINATION, §§ 5.13–5.17 (2d ed. 2017–2018).

110. 29 C.F.R. § 1607.5B (cross-references omitted).

111. 29 C.F.R. § 1607.16(D) ("Demonstrated by data showing that the content of a selection procedure is representative of important aspects of performance on the job.") Although "content validation" may seem to imply that the test in question is valid because it consists of samples of the job in question, the Guidelines require only that the employer demonstrate "data showing that the content of a selection procedure is representative of important aspects of performance on the job." 29 C.F.R. § 1607.16(D). Many tests have been found content-valid even though they do not, literally, consist of doing tasks required on the job. *See* Ass'n of Mexican-American Educators v. California, 183 F.3d 1055 (9th Cir. 1999) (the California teaching credential exam was content valid because the skills the test addressed were necessary for teachers, even though there was no showing that the test was in any way a sample of the job of teaching).

112. 29 C.F.R. § 1607.16(E) ("Demonstrated by data showing that the selection procedure measures the degree to which candidates have identifiable characteristics which have been determined to be important for successful job performance.")

113. 29 C.F.R. § 1607.15(B), defined as a "detailed statement of work behaviors and other information relevant to the job." 24 C.F.R. § 1607.16(K). Arti's method of proceeding will not qualify under traditional notions of that concept. *See* King & Mrkonich, *supra* note 9, at 577 ("Big Data begins from the opposite perspective—it searches first for correlations. The algorithm is uninterested in what any employee actually does, so long as the employer can identify who does it well and who does it poorly.").

tions.[114] Indeed, it is worth stressing that "content validation," which is the most common variety of test validation in the workplace, requires showing that the content of the selection device "is *representative of important aspects of performance* on the job for which the candidates are to be evaluated."[115] That could scarcely be possible without a job analysis. Similarly, it would not be possible to successfully establish construct validity without first determining the characteristics "important in successful performance in the job for which the candidates are to be evaluated."

While job tenure might be viewed as one aspect of "job performance," it seems unlikely that it is the only criterion relevant to that question, much less than it is "representative" of "important aspects" of performance. As we've noted, it's easy to list a number of qualities of a good worker, with longevity being pretty far down on most lists. In short, both content and construct validation seem to be off the table.

But Arti never claimed that job tenure was the theoretically best measure of a good employee; it claimed merely that it was the only criterion that, from the data available, could be used to predict success on the job.[116] Were Arti to manifest itself outside of this thought experiment, of course, it would likely fashion a more complex algorithm, one that took into account a wider variety of factors. Nevertheless, the point remains that Arti's output is the best of all possible worlds, at least given the data available and the limits of its processing power.

According to both Grimmelmann *et al.* and King *et al.*, that is precisely the problem. Data analytics characteristically looks to factors that correlate with (and hence are thought to predict) the desired criterion even though they are not necessarily ones for which a convincing explanation can be offered. Some supporters of analytics celebrate this: "Professor Viktor Mayer-Schonberger and Kenneth Cukier, Data Editor at *The Econo-*

114. The Guidelines focus on predicting job performance, but one might ask whether "business necessity" might be met by a policy that has little to do with job performance, as by identifying cost savings entailed by not hiring workers with certain characteristics (at most correlated with but not explicitly linked to protected classes). The Guidelines have little to say about that, and they were developed at a time when "business necessity" and "job performance" were treated as synonymous. Griggs v. Duke Power Co., 401 U.S. 424, 431 (1971) ("The Act proscribes not only overt discrimination but also practices that are fair in form, but discriminatory in operation. The touchstone is business necessity. If an employment practice which operates to exclude Negroes cannot be shown to be related to job performance, the practice is prohibited."). The 1991 Civil Rights Act's codification of disparate impact uses both terms, 42 U.S.C. 2000e-2(k)(1)(A)(i) (requiring the employer to "demonstrate that the challenged practice is job related for the position in question and consistent with business necessity," suggesting they may be separate concepts, but it uses them conjunctively, which suggest both are required. *See* Susan S. Grover, *The Business Necessity Defense in Disparate Impact Discrimination Cases*, 30 GA. L. REV. 387 (1996).

115. § 1607.5(B) (emphasis added).

116. Speaking of Arti making claims means that I am treating Arti as more and more human as this piece goes on. Maybe anthropomorphism is inevitable in the artificial intelligence space.

mist, declare, 'Causality won't be discarded, but it is being knocked off its pedestal as the primary fountain of meaning. Big data turbocharges non-causal analyses, often replacing causal investigations.'"[117] In other words, correlation is not causation, but who cares? The value of Big Data is that it can discover correlations that are hidden from human perception and, indeed, perhaps beyond human explanation. The literature, for example, suggests that taste for a particular manga site might be a good predictor of job success for coders.[118] Indeed, the place-of-residence criterion may be one whose relationship to longer job tenure is actually more intuitive than any number of correlations that analytics might discover and deploy.

A doctrinal answer to this debate would seem to lie in the Uniform Guidelines' approach to the third validation strategy, criterion validation. While this technique has so far been largely ignored in the antidiscrimination sphere because of its expense,[119] people analytics might require revisiting it. As we've seen, the Guidelines define criterion-related validity as a demonstration "by empirical data showing that the selection procedure is predictive of or significantly correlated with important elements of work behavior,"[120] which would seem to include Arti's operations. While "important elements" might suggest the need for a job analysis, the Guidelines explicitly exempt criterion validation from that requirement.[121] However, the Guidelines also anticipate that criteria validation will be used only when the criterion in question is plainly job-relevant: "Certain criteria may be used without a full job analysis if the user can show the importance of the criteria to the particular employment context. These criteria include but are not limited to production rate, error rate, tardiness, absenteeism, and length of service."[122]

Adherence to the Guidelines might, for example, justify Arti's use of place of residence (to the extent that it correlates with "length of service") even if it did not justify use of criteria that do not meet Grimmelmann & Westreich's demand for a rational explanation (rather than a mere statistical correlation). It seems, however, unlikely that Arti's correlation itself "shows" the importance of the criteria, and that no further demonstration need be made to satisfy them. In this, the Guidelines seem to track Grimmelmann's demand for a rational explanation.

---

117. King & Mrkonich, *supra* note 9, at 555.

118. *Id.* at 559.

119. *See* Guardians Ass'n v. Civil Service Comm'n, 630 F.2d 79, 92 (2d Cir. 1980) ("Content validation is generally feasible while construct validation is frequently impossible" because it requires "a criterion-related study, . . . a demonstration from empirical data that the test successfully predicts job performance."); *see also* RAMONA L. PAETZOLD & STEVEN L. WILLBORN, THE STATISTICS OF DISCRIMINATION § 4.13 (2016).

120. § 1607.16(F).

121. §§ 1607.14(A) & (B)(3).

122. § 1607.14(B)(3).

Yet another aspect of the Guidelines may also bear on the question. Arti will employ its criteria only when they are statistically significant,[123] but statistical significance means at most that the correlation is sufficiently unlikely to be the result of chance.[124] While such a showing will undoubtedly be necessary for a job-relation defense, there is also the issue of whether the employer will be required to show "practical significance," which focuses on the magnitude of the correlation.[125] While the cases to date have focused mostly on whether the plaintiff must establish both statistical and practical significance for her initial showing of disparate impact,[126] some decisions require the defendant to not only establish that the correlation is statistically significant but also that it is large enough in magnitude to justify its use.

Correlation coefficients are measured from +1 to -1, and it seems likely that extremely low coefficients, even though positive, will be unacceptable.[127] These are, however, precisely the kinds of correlations likely to be detected (at statistically significant levels) by Big Data. Unless Arti is programmed to avoid using coefficients below a certain point, its algorithm will likely use factors that no industrial psychologist would find acceptable for test validation. Indeed, it is somewhat more complicated than this: it may be that the entire algorithm produces results with both statistical and practical significance although factors incorporated in the algorithm, viewed individually, do not have practical significance. This

123. The level of significance can be important here. Traditionally, employment testing cases have used the .05 level. § 1607.14(B)(5).

124. *See* RAMONA L. PAETZOLD & STEVEN L. WILLBORN, THE STATISTICS OF DISCRIMINATION, §§ 4.13, 5.12 (2017-18 ed.).

125. *See* Ensley Branch of NAACP v. Seibels, 616 F.2d 812, 818 n.15 (5th Cir. 1980).

126. For example, in *Jones v. City of Boston*, 752 F.3d 38 (1st Cir. 2014), plaintiffs challenged the police department's use of hair samples to test for illegal drug use. While 98 percent of blacks passed the test, 99 percent of whites did so. Despite the argument that the difference was not large enough to be cognizable, the First Circuit held this proof sufficient: the data was statistically significant, and there is no requirement that plaintiffs must also prove "practical significance," i.e., that the size of the disparity is large enough to matter. *See also* Stagi v. Amtrak, 391 F. App'x 133 (3d Cir. 2010) (a showing of "statistical significance" regarding the sex impact of the challenged rule sufficient to avoid summary judgment even absent a finding of "practical significance"). *But see* Apsley v. Boeing Co., 691 F.3d 1184 (10th Cir. 2012) (statistical significance not sufficient to avoid summary judgment in light of practical insignificance). *See generally* Kevin Tobia, Note, *Disparate Statistics*, 126 YALE L.J. 2382 (2017).

127. The Uniform Guidelines do not directly address practical significance with regard to test validation. *Cf.* § 1607.4(D)(discussing both statistical and practical significance in proving a prima facie case of disparate impact), but courts nevertheless have viewed practical significance as a requirement. *E.g.*, Hamer v. City of Atlanta, 872 F.2d 1521, 1525–26 (11th Cir. 1989) (for a test to be criterion-related, the employer must establish both practical significance and statistical significance, with the former being the degree to which test scores relate to job performance). *See also* Dickerson v. U.S. Steel Corp., 472 F. Supp. 1304, 1348 (E.D. Pa. 1978) ("a low coefficient, even though statistically significant, may indicate a low practical utility.").

takes us back to the question of whether the algorithm should be viewed as a single selection device or broken into its component parts.[128]

In any event, whether these kinds of questions are a correct exegesis of the Guidelines is another question, but perhaps not the critical one. Just as *Teamsters* did not anticipate Arti, neither did either the Guidelines or the courts reviewing traditional selection devices. More to the point, courts 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. Further, as we've seen, the Guidelines are not governing "law," which means that they cannot be counted on to trump the deployment of artificial intelligence in the employment arena. Despite the "great deference" mantra accorded the Uniform Guidelines, the Court has repeatedly refused to give "great deference" (or even any meaningful deference) to other EEOC regulations under Title VII,[129] and one might predict with reasonable confidence that the Uniform Guidelines will not be assumed to govern a situation that they never envisioned, such as Arti.[130] In short, it may well be that the use of people analytics will open a whole new chapter of antidiscrimination law, requiring a fresh assessment of the meaning of validity in the disparate impact context.

For the sake of completeness, the final prong of disparate impact analysis should be at least mentioned although it seems unlikely to have any more traction here than in other contexts. Even if Arti's employer

---

128. *See supra* note 90.

129. *E.g.*, Vance v. Ball State Univ., 570 U.S. 421, 431 (2013) ("We reject the nebulous definition of a 'supervisor' advocated in the EEOC Guidance and substantially adopted by several courts of appeals."); Univ. of Tex. Sw. Med. Ctr. v. Nassar, 570 U.S. 338, 360-61 (2013) (denying "*Skidmore* deference" to the EEOC's regulation providing that "motivating-factor" analysis applied to Title VII retaliation cases). Court deference to the EEOC should be greater under the Americans with Disabilities Act since Title I conferred substantive rule-making authority on the EEOC in § 12116.

130. Barocas & Selbst argue that *Ricci v. DeStefano*, 557 U.S. 557 (2009), implicitly rejected the Guidelines. *See generally* Charles A. Sullivan, Ricci v. DeStefano*: End of the Line or Just Another Turn on the Disparate Impact Road?*, 104 NW. U. L. REV. 411 (2010). Although they recognize that *Ricci* did not directly address the issue, they argue that the Court's grant of summary judgment for the plaintiffs, despite the failure of the employer to find the at-issue test valid, established that test validation is unnecessary. Barocas & Selbst, *supra* note 6, at 671. They argue, further, that regulation of data mining to reduce biased models might be barred by *Ricci*. Professor Kim disagrees. Kim, *supra* note 1, at 930–31 (Barocas and Selbst "argue that attempts to regulate data mining are problematic because diagnosing the impact of a model requires taking protected class characteristics into account. [But] the problem in *Ricci* was not that the City took action with an awareness of its racial impact, but that the action entailed adverse employment actions against identifiable persons. Merely being aware of the racial consequences of a selection process does not constitute disparate treatment. Similarly, an employer's efforts to understand the racial consequences of its processes in order to avoid bias does not violate Title VII."); *see also* Kim, *Auditing Algorithms*, *supra* note 77, at 197 ("nothing in *Ricci* prohibits revising an algorithm after discovering it has discriminatory effects").

does establish a business necessity defense, a plaintiff may still hold the employer liable if she can show both the existence of "an alternative employment practice" with a lesser impact and that the employer "refuses to adopt such alternative employment practice."[131] Given that the operations of Arti are likely to be extraordinarily complex, this poses severe practical problems for plaintiffs seeking to show that an alternative with lesser impact would equally meet the employer's goals. Indeed, if Arti has done its task correctly, arguably no such proof would be possible,[132] although one could imagine a plaintiff successfully adducing evidence of employers using different selection criteria with a lesser impact.[133] But that aside, the requirement that the employer "refuses" to adopt such a practice seems to limit this surrebuttal to situations where a party in litigation learns enough about the disputed practice to serve something very much like a demand letter on the employer, which then fails to adopt the proffered alternative.[134] Perhaps needless to say, this surrebuttal has proved generally unsuccessful to date and is even less likely to be a meaningful check on data mining that passes the business necessity standard.

In short, the current state of disparate impact law leaves the legality if Arti's operations unclear. At most, its use of explicit classifiers on prohibited grounds would be barred under a pure causal analysis, but its achieving much the same result by relying on factors correlated with but not formally race or sex may well be permitted. Avoiding that result might be achieved by a focus on the technical aspects of Uniform Guidelines validation, but not only are they a slender legal reed but they are also profoundly unsatisfying from a normative perspective in this context.

IV. CONCLUSION

Arti may have taught us something, both about machine decision-making and human decision-making. One critical lesson is that the Supreme Court's bifurcation of all of discrimination into two theories, is very problematic, at least unless the theories are radically revised.

---

131. § 703(k)(1)(A)(ii).

132. King & Mrkonich, *supra* note 9, at 579–81.

133. The Supreme Court showed little sympathy for such an approach in *Ricci v. DeStefano*, 557 U.S. 557 (2009), where the City of New Haven defended its invalidation of a firefighter examination with a disparate impact in part by pointing to other fire departments that used different selection devices with a lesser impact. However, it may be that the problem was less the theory than the evidence adduced to support it. *Id.* at 591 ("[R]espondents refer to statements by Hornick in his telephone interview with the CSB regarding alternatives to the written examinations [such as 'assessment centers']. But Hornick's brief mention of alternative testing methods, standing alone, does not raise a genuine issue of material fact that assessment centers were available to the City at the time of the examinations and that they would have produced less adverse impact.").

134. This was essentially the strategy in one of the few successful invocations of the surrebuttal. Jones v. City of Boston, 845 F.3d 28 (1st Cir. 2016).

First, one may question a structure built on a vision of the human mind as easily capable of consciously separating out various inputs and able to avoid those that the law has prohibited. Reliance on those traits may be sufficient for liability but should not be necessary. Instead, the language of the statute suggests that a victim of Arti's sex-explicit sorting can bring suit because she has been denied a job "because of" her sex. That would mean not only that Arti could violate the statute by adapting explicit prohibited trait criteria but also, and important in the current world, that less conscious human motives could be sufficient.

As for disparate impact, the Arti experiment underscores the problems of deploying traditional analysis to assess a technological breakthrough. Indeed, there is a striking parallel between the invention of disparate impact in *Griggs v. Duke Power*[135] and the potential uses of data analytics in employee selection. In *Griggs*, the Court dealt with the then-common use of testing by public and private employees, which created an enormous structural barrier to the advancement of African Americans. It noted that "[h]istory is filled with examples of men and women who rendered highly effective performance without the conventional badges of accomplishment in terms of certificates, diplomas, or degrees. Diplomas and tests are useful servants, but Congress has mandated the common-sense proposition that they are not to become masters of reality."[136] Today, we may be on the verge of reconstructing such structural barriers under the flag of Big Data.

Whether disparate impact theory is well equipped to answer the questions posed by these new developments remains to be seen. Is correlation between some criterion and some measure of job performance enough? Is an additional showing necessary that the measure is "representative" or "important"? As Grimmelmann and Westreich argue, is a plausible causal explanation required?

Arti, alas, can't answer these questions. It's up to mere humans.

---

135. 401 U.S. 424 (1971).

136. *Id.* at 433.