



**When You Gaze Into the Abyss,
the Abyss Gazes Into You-**

**The Unintended Consequences
of Data Analytics**

Michael T. Anderson
Murphy Anderson PLLC
111 Devonshire St. 5th Flr.
Boston, MA 02109
617-227-5720
manderson@murphypllc.com

Big Data is a double-edged tool. Data mining gives employers a sophisticated analysis of information about applicants and employees. But Big Data also makes the employer's thinking more transparent. Plaintiffs, workers and unions will now be able to probe more precisely what criteria the employer is using to evaluate the workforce. This allows litigants to frame theories of liability about the employer's search criteria, which had been clouded by employers' more subjective discretion in the past.

This matters in litigation. Just as employers are now able to analyze applicant and employee pools with unprecedented quantitative power, plaintiffs in employment litigation now have a robust tool to prove that the employer's decisions are linked to legally protected categories.

This will show up in one of two related areas:

Identity

Search criteria may have a disproportionate effect on protected status.

Where the law protects identity characteristics like race, sex, or age, employers may face unexpected liability where search criteria disproportionately affect the protected classification.

This “disparate impact” theory of liability has been the law since the 1970s. See 42 U.S.C. §2000e-2(a) and (k). Big Data multiplies the number of objectively provable criteria that may have a disparate impact.

Behavior

Big Data may also produce discrimination against behavior and attitudes correlated with protected categories. While EEO law has historically focused on identity, a growing area of concern involves employee behavior resisting illegal practices. This is the “opposition” clause in Title VII, 42 U.S.C. §2000e-3(a), which mirrors anti-retaliation provisions in the FLSA, 29 U.S.C. §215(a)(3) and the National Labor Relations Act, 29 U.S.C. §158(a)(1) and (3). As plaintiffs discover what behavior is favored or disfavored in Big Data analysis, there will be growing pressure to expand the boundaries of protected behavior to prevent employers from weeding out workers who are likely to oppose employer violations.

I. Disparate Impact on Identity

A. Historic framework

The law has known the “disparate impact” theory since the 1970s. Under that theory, facially neutral employment policies that are “fair in form,” *i.e.*, they apply equally to blacks and whites or to women and men, nonetheless may be deemed unlawful if they are “discriminatory in operation.” *Griggs v. Duke Power Co.*, 401 U.S. 424, 431 (1971); *see, e.g., Dothard v. Rawlinson*, 433 U.S. 321, 327 (1977) (disparate impact on women of uniformly applied height and weight restrictions established a *prima facie* case of discrimination).

In 1991, Congress codified the theory in 42 U.S.C. §2000e-2(k):

(k) Burden of proof in disparate impact cases

(1)(A) An unlawful employment practice based on disparate impact is established under this subchapter 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; or

(ii) the complaining party makes the demonstration described in subparagraph (C) with respect to an alternative employment practice and the respondent refuses to adopt such alternative employment practice.

(B) (I) With respect to demonstrating that a particular employment practice causes a disparate impact as described in subparagraph

(A)(I), the complaining party shall demonstrate that each particular challenged employment practice causes a disparate impact, except that if the complaining party can demonstrate to the court that the elements of a respondent's decisionmaking process are not capable of separation for analysis, the decisionmaking process may be analyzed as one employment practice.

(ii) If the respondent demonstrates that a specific employment practice does not cause the disparate impact, the respondent shall not be required to demonstrate that such practice is required by business necessity.

* * *

(2) 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 subchapter.

B. How Big Data accelerates plaintiffs' cases

1. Correlations with protected status

The disparate-impact theory in the 1970s typically focused on direct, open job requirements like a high-school education (deemed insufficiently related to blue-collar work in *Griggs v. Duke Power Co.*, 401 U.S. 424, 431 (1971) to justify disparate impact on race in hiring), or height and weight requirements (deemed insufficiently related to prison guard work in *Dothard v. Rawlinson*, 433 U.S. 321 (1977) to justify disparate impact on women).

The EEOC's regulations governing statistical analysis, 29 C.F.R. §§ 1607.01-.14, were promulgated in 1978. They reflect the more rudimentary status of statistical testing in the 1970s. Users of Big Data now will immediately understand the unhelpful vagueness of standards like that

outlined in 29 C.F.R. § 1607.14(C)(5):

To demonstrate the content validity of a selection procedure, a user should show that the behavior(s) demonstrated in the selection procedure are a representative sample of the behavior(s) of the job in question or that the selection procedure provides a representative sample of the work product of the job. . . . The closer the content and the context of the selection procedure are to work samples or work behaviors, the stronger is the basis for showing content validity. As the content of the selection procedure less resembles a work behavior, or the setting and manner of the administration of the selection procedure less resemble the work situation, or the result less resembles a work product, the less likely the selection procedure is to be content valid, and the greater the need for other evidence of validity.

With Big Data, it is much harder to apply the EEOC regulations, because analytics now measure many more times as many search criteria, and follow more complicated algorithms, than the testing contemplated in 1978.

This raises the stakes, because Big Data makes employer decision-making depend on a larger number of search criteria. For example, an employer metric that evaluates workers, *inter alia*, by multiple factors including their zip code will be open to a charge of red-lining the workforce based on race or national origin because it identifies residence in non-white neighborhoods as a negative trait.

The danger that a jury may seize on any one metric means that the employer may have to abandon any Big Data analysis that does its stated job -- to distinguish among applicants in meaningful ways.

2. Difficulty in claiming innocent intent

Before Big Data, employer classifications were typically simple and publicly stated. Even where policies like the high-school diploma requirement in *Griggs* ran afoul of disparate-impact doctrine, the openness with which they were maintained tended to preclude any claim of intentional discrimination. As a result, disparate-impact claims have been much more easily met with *bona fide* business justifications for any differential effects.

But Big Data tends to conceal the employer's thinking, at least until discovery. Employers using data mining may not even be aware of how their algorithms are processing the data. Once the search criteria are discovered, however, the hidden nature of their operation makes a charge of pretext more plausible. Since 42 U.S.C. §2000e-2(k)(2) deprives employers of a "business justification" defense for intentional discrimination, litigation will focus on whether employers understood that their search criteria might skew the workforce. If a jury accepts that inference, the effects of data mining will not be insulated by any *bona fide* business justification for the search criteria.

As a result, legal protection for categories like race, sex, age and national origin will spread to many, if not most, population characteristics that are the normal focus of Big Data analysis.

II. Disparate Impact on Protected Behavior

A. Big Data may screen out workers likely to oppose management.

EEO law is gradually migrating toward labor law, which understands worker resistance to management authority as “protected concerted activity.” This is a behavioral trait protected even as to applicants. For example, the NLRB has applied disparate-impact theory in striking down facially neutral employer policies that tend to select out union activists. In *Aztech Electric Co.*, 335 NLRB 260, 263 (2001), the NLRB held that a non-union employer’s policy against hiring applicants who had previously earned more than 30% than its own wage was an unlawful screen to keep union members out.

The NLRB will be called on to address more sophisticated forms of anti-union screening through Big Data. For example, if a large employer like Walmart can be shown to screen applicants to disfavor behavioral traits like independent thinking, resistance to authority, or loyalty to fellow workers, the NLRB will eventually have to identify these characteristics as protected.

B. Employers’ common-law assumptions easily run afoul of statutory protections.

The problem for employers is that anti-discrimination law (*e.g.*, the NLRA, FLSA and Title VII), by definition, prohibits employers from considering some matters that the common law broadly allowed them to

consider. Why shouldn't an employer want to hire only employees with a high school diploma, *cf. Griggs*, or employees from more stable neighborhoods, or employees whose profile suggests higher productivity?

The short answer is that many common-law assumptions about what an employee may lawfully consider are no longer valid. For example, the NLRB has consistently struck down language in employee handbooks that most employers had considered basic. Rules requiring employees to keep their wages and job conditions confidential, *Cintas Corp.*, 344 NLRB 943, 943 (2005) *enfd.* 482 F.3d 463 (D.C. Cir. 2007); requiring loyalty to management and forbidding "derogatory behavior," *Southern Maryland Hospital Center*, 293 NLRB 1209, 1222 (1989) *enfd.* 916 F.2d 932 (4th Cir. 1990); or wearing controversial buttons on work uniforms. *Pioneer Hotel*, 324 NLRB 918, 923 (1997) *enfd.in relevant part*, 182 F.3d 939, 946 (D.C. Cir. 1999).

Like the authors of these offending employee handbooks, the designers of Big Data metrics may be completely unaware of just what the law does and does not prohibit.

C. The scope of opposition clauses in EEO and FLSA law will expand.

As currently written, the opposition clauses of Title VII, 42 U.S.C. §2000e-3(a), and the anti-retaliation provision in the Fair Labor Standards Act, 29 U.S.C. §215(a)(3), only prohibit to retaliation against a current

employee after he or she has taken a concrete step to oppose an unlawful practice. But even now, the boundaries of what constitutes “opposition” have been under strong pressure, with the courts holding that informal behavior, and even response to employer questions, is protected. *See, e.g., Crawford v. Metro. Government of Nashville*, 555 U.S. 271, 276 (2009) (Title VII anti-retaliation protection extends to employee who did not initiate protest, but merely answered employer questions); *Kasten v. Saint-Gobain Performance Plastics Corp.*, 563 U.S. 1, 8 (2011) (FLSA anti-retaliation clause applies to informal resistance to policy, not merely formal complaint).

Big Data will create pressure for courts and Congress to extend this protection to applicants and employees from discrimination because they are likely to oppose such practices. That is, the logic of the law would prohibit an employer from systematically refusing to hire applicants whose profile indicates behavior or attitudes associated with opposition.

This is already a reality in the National Labor Relations Act, 29 U.S.C. §158(a)(1) and (3), since an “employee” protected from discrimination includes an applicant for employment whose past union activity, which indicates likely future union activity if hired. *NLRB v. Town & Country Electric, Inc.*, 516 U.S. 85, 88 (1995), *citing Phelps Dodge Corp. v. NLRB*, 313 U.S. 177, 185-186(1941).It is also suggested in the Sarbanes-Oxley Act, 18 U.S.C.

§1514A, where the boundaries of whistleblower protection are not confined to employees of the public company itself, but also extend to whistleblowing by employees of private companies related to the public issuer. *Lawson v. FMR LLC*, ___ U.S. ___, 134 S.Ct. 1158, 1163 (2014). After *Lawson*, a public company could not safely refuse to hire an applicant solely because he or she had reported fraud in the past as to another company.

Discovery of employer search criteria will generate pressure to address employer analytics that purge the workforce of people most likely to oppose unlawful practices. For example, if an employer weeds out applicants because their profile suggests they are skeptical of authority, does this amount to a pretext for anti-union discrimination? May an employer avoid hiring applicants who have a history of litigation, where this may purge the workforce of employees most likely to invoke the opposition clause of anti-discrimination law? May an employer select among employees based on personal traits that suggest they are more likely to marry and have children, if that translates to prospective pregnancy discrimination?

These are issues that Big Data places in the forefront. The power of Big Data permits employers a profound insight into global data about applicants and employers; however, it invites equally intense scrutiny of the employer's own thinking and motivations as well.