

NOTES

GENDER-NORMED PHYSICAL-ABILITY TESTS UNDER TITLE VII

*Eve A. Levin**

Employers seeking to test job applicants for strength or speed while adhering to the mandates of Title VII often use gender-normed physical-ability tests. Gender-normed tests set different raw cutoffs for male and female applicants such that each class would be expected to have roughly equal pass rates. This practice has helped employers—especially law enforcement agencies—retain physical hiring standards while mitigating their disparate impact on women.

In Bauer v. Lynch, the Fourth Circuit became the first court of appeals to directly consider the permissibility of gender-norming physical-ability tests under Title VII. Though the court concluded that the gender-norming did not itself constitute a form of discrimination, to do so, it applied the so-called unequal-burdens test, a much-maligned doctrine that had previously been applied in only one area of Title VII jurisprudence: appearance and grooming standards.

This Note reexamines the practice of gender-norming physical-ability tests in light of the Bauer decision. It argues that contrary to the Bauer court's decision, gender-norming physical-ability tests is a form of discrimination under Title VII that must be justified by a bona fide occupational qualification and that neither the unequal burdens doctrine nor any other Title VII carveout can excuse employers from carrying that burden. It then provides a normative defense of that doctrinal conclusion. While unitary hiring standards that impose a disparate impact on women perpetuate the gender hierarchy by exclusion, job-unrelated, gender-normed physical-ability tests perpetuate the gender hierarchy by arbitrarily privileging masculinity while evading judicial review. Thus, applying a demanding business justification for physical-ability tests—the bona fide occupational qualification for gender-normed tests and the job-relatedness–business-necessity standard for tests with a disparate impact—best serves Title VII's antidisubordination principle.

INTRODUCTION

The representation of women in law enforcement agencies nationwide has improved since Congress enacted the Civil Rights Act of 1964,

* J.D. Candidate 2018, Columbia Law School.

in a few cases drastically,¹ but the progress has been uneven and marred by episodes of intentional discrimination.² Changes in the way law enforcement organizations screen applicants for physical ability have been a big part of the story of increased representation in law enforcement,³ but even newer physical-ability tests (PATs) have faced substantial scrutiny in the courts under Title VII's disparate impact prohibition.⁴

To avoid incurring disparate impact liability, some law enforcement agencies use gender-normed PATs.⁵ Gender-normed tests use different

1. See Lynn Langton, Bureau of Justice Statistics, U.S. Dep't of Justice, Women in Law Enforcement, 1987–2008, *passim* (2010), <http://www.bjs.gov/content/pub/pdf/wle8708.pdf> [<http://perma.cc/CQ77-8FP9>] (indicating modest aggregate improvements in gender diversity in law enforcement); Brian A. Reaves, Bureau of Justice Statistics, U.S. Dep't of Justice, Local Police Departments, 2013: Personnel, Policies, and Practices 4 (2015), <http://www.bjs.gov/content/pub/pdf/lpd13ppp.pdf> [<http://perma.cc/H6V4-B33K>] (noting an increase in female representation in local police departments since 1987).

2. David Alan Sklansky, Not Your Father's Police Department: Making Sense of the New Demographics of Law Enforcement, 96 J. Crim. L. & Criminology 1209, 1213 (2006) (“How have the demographics of American police departments changed since the 1960s? The short answer is by quite a lot, although not as much as might be hoped, and at a widely varying pace.”); see also Civil Rights Div., U.S. Dep't of Justice & EEOC, Diversity in Law Enforcement: A Literature Review 1 (2015), https://cops.usdoj.gov/pdf/taskforce/Diversity_in_Law_Enforcement_Literature_Review.pdf [<http://perma.cc/2FH4-VDDU>] (“Unfortunately, intentional employment discrimination still remains a substantial barrier in the law enforcement context.”). This truism applies to gender, race, and sexual-orientation diversity. See Sklansky, *supra*, at 1213 (noting that, though minority representation within American police departments has improved since the 1960s, diversity with regard to gender, race, and sexual orientation has not increased “as much as might be hoped”). Related public safety roles have had similar troubles. Cf. Ricci v. DeStefano, 557 U.S. 557, 609 (2009) (Ginsburg, J., dissenting) (“Firefighting is a profession in which the legacy of racial discrimination casts an especially long shadow.”).

3. See Gregory S. Anderson et al., Police Officer Physical Ability Testing: Re-Validating a Selection Criterion, 24 Policing 8, 9 (2001) (recounting the transition from height and weight requirements to less discriminatory measures over the past half-century); Michael L. Birzer & Delores E. Craig, Gender Differences in Police Physical Ability Test Performance, 15 Am. J. Police 93, 93 (1996) (noting the changes over time in the use of physical-ability tests and the effect on gender balance in police forces).

4. See 42 U.S.C. § 2000e-2(k) (2012); see also Ernst v. City of Chicago, 837 F.3d 788, 795–805 (7th Cir. 2016) (invalidating a PAT for fire department paramedics because of a disparate impact on women); Pietras v. Bd. of Fire Comm'rs, 180 F.3d 468, 474–75 (2d Cir. 1999) (upholding a district court's finding that a firehose-dragging exercise for firefighting applicants had an impermissible disparate impact on women); Harless v. Duck, 619 F.2d 611, 616 (6th Cir. 1980) (invalidating a PAT for police applicants on a disparate impact theory); Yiyang Wu, Comment, Scaling the Wall and Running the Mile: The Role of Physical-Selection Procedures in the Disparate Impact Narrative, 160 U. Pa. L. Rev. 1195, 1212 n.82 (2012) (collecting disparate impact PAT cases). Similar challenges have also arisen outside the law enforcement and public safety contexts. See, e.g., EEOC v. Dial Corp., 469 F.3d 735, 742–43 (8th Cir. 2006) (sustaining a disparate impact challenge to a PAT for factory workers).

5. Kimberly A. Lonsway, Nat'l Ctr. for Women & Policing, Tearing Down the Wall: Problems with Consistency, Validity, and Adverse Impact of Physical Agility Testing in Police Selection, 6 Police Q. 237, 258 (2003) [hereinafter Lonsway, Tearing Down the Wall] (finding that between one-quarter and one-third of police departments that use

cutoff scores for male and female applicants, such that men and women would be expected to pass at equal rates. In other words, a unitary standard would apply the same cutoff to all applicants—say, twenty push-ups for all—but a gender-normed standard would apply different raw cutoffs to men and women—say, fourteen push-ups for women and thirty for men.⁶ Of the law enforcement agencies administered by the Department of Justice (DOJ), at least four use gender-normed physical-ability or strength cutoffs for employment.⁷ At times, courts considering disparate impact claims have blessed gender-norming as a permissible way to retain physical selection devices while adhering to the mandates of antidiscrimination law.⁸ An October 2016 report issued jointly by the DOJ and the Equal Employment Opportunity Commission (EEOC) cited

PATs gender-norm those tests). To the author's knowledge, there are no more recent statistical studies of the use of gender-normed PATs, and the National Center for Women and Policing, the organization that conducted the cited 2003 study, has been defunct since at least 2013 due to funding shortages. See Jay Newton-Small, *There Is a Simple Solution to America's Policing Problem: More Female Cops*, *Time* (July 14, 2016), <http://time.com/4406327/police-shootings-women-female-cops/> [<http://perma.cc/V5U8-ZFC5>].

6. See *infra* section II.A (describing the gender-normed PAT used by the FBI).

7. See U.S. DEA Training Acad., *DEA Basic Agent Training: Physical Training & Conditioning Manual* 50–51, <http://www.dea.gov/careers/agent/DEA%20Basic%20Agent%20Training%20-%20Physical%20Fitness%20Manual%20PTT%20Protocols.pdf> [<http://perma.cc/LQG2-GRPN>] (last visited Nov. 1, 2017) (detailing the gender-normed PAT for the Special Agent program in the Drug Enforcement Administration); Pre-Employment Physical Task Test, Bureau of Alcohol, Tobacco, Firearms & Explosives, <https://www.atf.gov/careers/pre-employment-physical-task-test> [<http://perma.cc/R85E-5AQ3>] (last visited Oct. 13, 2017) (describing the gender-normed and age-normed PAT for applicants to the special agent program in the Bureau of Alcohol, Tobacco, Firearms and Explosives); *infra* section II.A (describing the gender-normed Physical Fitness Test for applicants to the FBI's Special Agent program). Compare Fitness Standards for Women, U.S. Marshals Serv., http://www.usmarshals.gov/careers/fitness_women.html [<http://perma.cc/AG6D-J85D>] (last visited Oct. 13, 2017) (describing fitness requirements for female applicants to the U.S. Marshals Service), with Fitness Standards for Men, U.S. Marshals Serv., http://www.usmarshals.gov/careers/fitness_men.html [<http://perma.cc/23F9-KWV6>] (last visited Oct. 13, 2017) (describing the same for male applicants). By contrast, the Bureau of Prisons requires new hires to pass a unitary PAT designed to measure employees' "ability to perform the essential functions of a correctional worker," including, *inter alia*, a dummy drag, "self-defense movements," and timed skill components (for example, running a quarter-mile and applying handcuffs within a certain time period). Our Hiring Process, Fed. Bureau of Prisons, http://www.bop.gov/jobs/hiring_process.jsp [<http://perma.cc/7BM2-XFGM>] (last visited Oct. 13, 2017).

8. See, e.g., *Lanning v. Se. Pa. Transp. Auth.*, 181 F.3d 478, 490 n.15 (3d Cir. 1999) (making several recommendations to the Southeastern Pennsylvania Transportation Authority, including that it "institute a non-discriminatory test . . . such as a test that would exclude 80% of men as well as 80% of women through separate aerobic capacity cutoffs for the different sexes"); cf. *United States v. Virginia (VMI)*, 518 U.S. 515, 550 n.19 (1996) (observing in the equal protection context that "[a]dmitting women to [the Virginia Military Institute] would undoubtedly require alterations necessary . . . to adjust aspects of the physical training programs" but that "[e]xperience shows such adjustments are manageable").

the practice approvingly as a means for law enforcement agencies to mitigate the disparate impact of PATs.⁹

In *Bauer v. Lynch*,¹⁰ the Fourth Circuit became the first court of appeals to directly consider the permissibility of gender-norming PATs under Title VII.¹¹ The court concluded that the gender-norming did not itself constitute a form of discrimination. But to do so, it applied the so-called unequal-burdens test, a much-maligned doctrine that had been applied in only one area of Title VII jurisprudence: appearance and grooming standards.¹² The doctrine as previously understood is at best an uncomfortable fit with the facts of the *Bauer* case.¹³ Further, Title VII's plain text explicitly prohibits adjusting cutoff scores on the basis of race, sex, color, or national origin, seemingly at odds with the practice of gender-norming these tests.¹⁴

The *Bauer* decision occasions a reexamination of this practice. Part I of this Note surveys the relevant legal backdrop, beginning with the Title VII disparate impact framework and challenges to PATs under that theory. Part II explores whether, as a descriptive matter, gender-normed PATs are permitted under Title VII. The Fourth Circuit's *Bauer* opinion provides the jumping-off point for this discussion. This Part concludes that the *Bauer* court got it wrong and that Title VII does not permit gender-norming absent a valid business justification. Part III provides a normative defense of the doctrinal conclusion reached in Part II. It argues that courts should reject the Fourth Circuit's reasoning and instead apply a traditional Title VII disparate treatment analysis to gender-normed PATs, requiring the norming, as a distinction based on sex, to be justified as a bona fide occupational qualification. Such an approach would, perhaps counterintuitively, better promote gender justice in the workplace.

9. U.S. Dep't of Justice & EEOC, *Advancing Diversity in Law Enforcement* 55 (2016), <http://www.justice.gov/crt/case-document/file/900761/download> [<http://perma.cc/N3DZ-R55F>].

10. 812 F.3d 340 (4th Cir.), cert. denied, 137 S. Ct. 372 (2016).

11. The question has been adjudicated three other times: once by a federal district court, once by a Vermont state court, and once by an administrative law judge in an EEOC proceeding. Each concluded that gender-normed PATs are not disparate treatment under Title VII. See *Powell v. Reno*, No. 96-2743 (NHJ), 1997 U.S. Dist. LEXIS 24169, at *11 (D.D.C. July 24, 1997); *In re Scott*, 779 A.2d 655, 661 (Vt. 2001); *Hale v. Holder*, No. 570-2007-00423X (E.E.O.C. Sept. 20, 2010); see also *Alspaugh v. Comm'n on Law Enf't Standards*, 634 N.W.2d 161, 169 (Mich. Ct. App. 2001) (using Title VII principles as a guide to assessing gender-normed PATs under the Michigan Civil Rights Act and holding that the PATs were valid).

12. See *infra* section II.B.2. For cases applying the unequal-burdens doctrine in the grooming and appearance context, see *infra* note 101.

13. See *infra* notes 112–114, 134–138 and accompanying text (arguing that the unequal-burdens doctrine's traditional justifications do not apply to the context of gender-normed fitness standards).

14. 42 U.S.C. § 2000e-2(*l*) (2012).

I. THE DISPARATE IMPACT OF PHYSICAL-ABILITY TESTS

Title VII proscribes two basic forms of discrimination: disparate treatment—decisions and policies that intentionally discriminate on the basis of a protected characteristic—and disparate impact—neutral policies that have discriminatory effects.¹⁵ An employer seeking to use a PAT must ensure the test avoids both pitfalls or else provide a business justification. It is the latter theory of discrimination that has proven the more troublesome hurdle for PATs. Section I.A therefore outlines Title VII's disparate impact protections, and section I.B reviews how courts have applied this theory to PATs. In sum, Part I describes the legal framework that provides the impetus for employers to adopt the kinds of practices challenged in Parts II and III.

A. *Disparate Impact Challenges: The Basic Framework*

Almost all Title VII challenges to PATs have been disparate impact claims. Most of these cases don't involve class-normed tests but rather unitary requirements—single cutoffs that apply, without adjustment, to all test takers.¹⁶ The main concern of this Note is gender-normed testing that, by definition, should not have a disparate impact on women.¹⁷ Nonetheless, understanding the relative success of disparate impact challenges to PATs is crucial to understanding both why employers adopt gender-normed tests in the first place and the alternatives available to them.

The Supreme Court first endorsed the availability of a disparate impact cause of action under Title VII in its landmark decision *Griggs v. Duke Power Co.*¹⁸ The employment practice at issue in *Griggs* was a facially neutral education requirement that operated to exclude the vast majority

15. 1 Merrick T. Rossein, *Employment Discrimination Law and Litigation* § 2:1 (2016).

16. See, e.g., Wu, *supra* note 4, at 1212–28 (arguing that physical-selection cases have been a uniquely successful subset of sex-based disparate impact claims).

17. Though gender-normed tests attempt to mitigate inter-class performance gaps, in practice they often fail to completely do so. Several disparate impact cases have challenged gender-normed tests on the basis that their effects still worked to exclude a disproportionate share of female applicants, norming notwithstanding. In other words, in distinction to *Bauer v. Lynch* and other cases challenging gender-norming as such under a disparate treatment theory, the disparate impact cases have challenged PATs basically because they aren't gender-normed *enough*. Compare *Easterling v. Connecticut*, 783 F. Supp. 2d 323, 325–36 (D. Conn. 2011) (challenging a 1.5-mile-run test for correctional officer applicants that set different cutoff scores for male and female applicants under a disparate impact theory), and *United States v. City of Erie*, 411 F. Supp. 2d 524, 528–29 (W.D. Pa. 2005) (challenging a test for police candidates that required male applicants to complete more pull-ups than female applicants under a disparate impact theory), with *Bauer v. Lynch*, 812 F.3d 340, 346 (4th Cir.) (challenging a gender-normed PAT under a disparate treatment theory), cert. denied, 137 S. Ct. 372 (2016).

18. 401 U.S. 424, 431 (1971).

of black workers from desirable promotions and roles.¹⁹ The Court, acknowledging that the practice may not have been intentionally discriminatory, famously pronounced that “good intent or absence of discriminatory intent does not redeem employment procedures or testing mechanisms that operate as ‘built-in headwinds’ for minority groups and are unrelated to measuring job capability.”²⁰ The Court concluded: “Nothing in the Act precludes the use of testing or measuring procedures; obviously they are useful. What Congress has forbidden is giving these devices and mechanisms controlling force unless they are demonstrably a reasonable measure of job performance.”²¹ Thus, facially neutral practices—even those not motivated by discriminatory intent or animus—may nonetheless contravene Title VII if, in effect, they work to arbitrarily exclude protected classes at disproportionate rates.²² Congress subsequently codified the disparate impact framework in the Civil Rights Act of 1991.²³

To make out a *prima facie* case of disparate impact, a plaintiff must show an employer uses a “particular employment practice that causes a disparate impact on the basis of race, color, religion, sex, or national origin.”²⁴ Often, disparate impact plaintiffs challenging selection devices make out a *prima facie* case by satisfying the four-fifths rule;²⁵ that is, by showing with statistical evidence that the plaintiff’s protected class has a pass rate that is at most four-fifths that of the favored class.²⁶ Upon such a

19. *Id.* at 426–29.

20. *Id.* at 432.

21. *Id.* at 436.

22. Disparate treatment challenges brought by favored-group members, like Jay Bauer’s claim, are fairly routine. But the possible application of the disparate impact framework to this sort of claim by a favored-group member poses distinctive problems. For an extensive discussion of this question, see generally Charles A. Sullivan, *The World Turned Upside Down?: Disparate Impact Claims by White Males*, 98 *Nw. U. L. Rev.* 1505 (2004).

23. Civil Rights Act of 1991, Pub. L. No. 102-166, § 105, 105 Stat. 1071, 1074–75 (codified as amended at 42 U.S.C. § 2000e-2(k) (2012)). The need for the statutory amendment arose after a series of Supreme Court decisions significantly walked back *Griggs*’s reach. See *Wards Cove Packing Co. v. Atonio*, 490 U.S. 642, 660 (1989) (holding that an employer has a burden of only production, rather than persuasion, on the question of business necessity); *Watson v. Fort Worth Bank & Tr.*, 487 U.S. 977, 998 (1988) (deciding in favor of a disparate impact plaintiff while nonetheless requiring, for purposes of the affirmative “business necessity” defense, that the defendant show only some *generalized* relationship between the challenged selection device and the job).

24. 42 U.S.C. § 2000e-2(k) (1) (A) (i).

25. See, e.g., *Watson*, 487 U.S. at 995 n.3 (describing the use of the four-fifths rule by lower courts).

26. See EEOC Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. § 1607.4(D) (2017) (“A selection rate . . . which is less than four-fifths . . . of the rate for the group with the highest rate will generally be regarded . . . as evidence of adverse impact, while a greater than four-fifths rate will generally not be regarded . . . as evidence of adverse impact.”). For criticism of the four-fifths rule, see generally Jennifer L. Peresie,

showing, the burden of persuasion shifts to the employer to show “that the challenged practice is job related for the position in question and consistent with business necessity.”²⁷ If the employer does so, the plaintiff can prevail only by showing there exists an alternative practice “that has less disparate impact and serves the employer’s legitimate needs.”²⁸

The only Supreme Court case to analyze a sex discrimination claim under the disparate impact framework is *Dothard v. Rawlinson*.²⁹ In *Dothard*, the female plaintiff challenged the Alabama Board of Corrections’s height and weight requirements for correctional officer positions.³⁰ The combination of the height and weight requirements would have excluded 41.13% of the female population of the United States; it would have excluded less than 1% of the male population.³¹ Based on the dramatic statistical difference, the Court found that the plaintiff had made out a prima facie case of disparate impact.³²

The defendants responded that height and weight requirements “have a relationship to strength, a sufficient but unspecified amount of which is essential to effective job performance as a correctional counselor.”³³ Therefore, the defendants argued, the requirements were sufficiently job related and consistent with business necessity to rebut the prima facie case.³⁴ The Court rejected that argument as well, noting that the defendants “produced no evidence correlating the height and weight requirements with the requisite amount of strength thought essential to good job performance.”³⁵ The Court also observed that “[i]f the job-related quality that the appellants identify is bona fide, their purpose

Toward a Coherent Test for Disparate Impact Discrimination, 84 Ind. L.J. 773 (2009) (arguing that the four-fifths rule is an insufficient measure of disparate impact causation and should be combined with a test for statistical significance); Elaine W. Shoben, Differential Pass-Fail Rates in Employment Testing: Statistical Proof Under Title VII, 91 Harv. L. Rev. 793, 805–06 (1978) (“The four-fifths rule is an ill-conceived resolution It will produce anomalous results in certain cases because it fails to take account of differences in sampling size. It also neglects the magnitude of differences in pass rates by considering only the ratio of the two rates.”).

27. 42 U.S.C. § 2000e-2(k)(1)(A)(i).

28. *Ricci v. DeStefano*, 557 U.S. 557, 578 (2009) (citing 42 U.S.C. § 2000e-2(k)(1)(A)(ii), (C)). For a PAT disparate impact case emphasizing the importance of the third, alternative-practice prong, see *United States v. Massachusetts*, 781 F. Supp. 2d 1, 20 (D. Mass. 2011) (“The real question is not whether the [challenged] PAT results in a disparate impact on women (it does), nor is it whether the test is job related and implemented to achieve public policy goals (we’ll see). The real question is whether we can do better.”).

29. 433 U.S. 321 (1977).

30. *Id.* at 323–24.

31. *Id.* at 329–30.

32. *Id.* at 331.

33. *Id.*

34. *Id.*

35. *Id.*

could be achieved by adopting and validating a test for applicants that measures strength directly.”³⁶ This, the Court said, “would fully satisfy the standards of Title VII because it would be one that ‘measure[d] the person for the job and not the person in the abstract.’”³⁷ Thus, the Court’s determination appeared to turn on three deficiencies: first, a lack of evidence tying the selection device (height and weight) to the desired quality (some unspecified amount of strength); second, a lack of evidence showing that the unspecified amount of strength was job related with respect to the corrections officer position; and third, skepticism that the nexus between strength and the selection device was sufficiently close, as reflected in the Court’s suggestion that the defendant adopt a test that “measure[d] strength directly.” This final point could be read as doubt either that the height and weight requirements could possibly be validated as a legitimate way to measure strength or that even if they were, a future plaintiff would nonetheless be able to show that there exist less discriminatory alternative practices that adequately measure strength.³⁸

B. *Disparate Impact Challenges to Physical-Ability Tests*

Litigants³⁹ have had relative success in the lower courts challenging PATs under Title VII’s disparate impact prohibition.⁴⁰ Because statistical evidence supporting a prima facie case is usually forthcoming in these cases, the outcome often turns on the application of the job-relatedness—

36. *Id.* at 332.

37. *Id.* (quoting *Griggs v. Duke Power Co.*, 401 U.S. 424, 436 (1971)).

38. The Court went on to address and ultimately reject the plaintiff’s other claim. The same statutory scheme that imposed the height and weight requirements also reserved certain positions designated as “contact positions” to men. Though this provision explicitly discriminated on the basis of sex, the Court nonetheless found that being a man was a bona fide occupational qualification for a contact position, and therefore excused the employer’s disparate treatment. *Id.* at 332–37.

39. These cases have overwhelmingly been sex discrimination claims, though some data suggest that unitary physical standards can also disadvantage applicants along racial, ethnic, and national-origin lines. See *Birzer & Craig*, *supra* note 3, at 94 (collecting studies showing the adverse impact of physical-ability testing on Asian and Hispanic applicants to law enforcement positions). Several challenges to the older tests, especially those tests that relied on crude height and weight requirements, were brought and sustained as race or national-origin discrimination. See, e.g., *United States v. City of Buffalo*, 457 F. Supp. 612, 625 (W.D.N.Y. 1978) (invalidating physical hiring requirements, including a height requirement, on the basis of a disparate impact on Hispanic applicants); *League of United Latin Am. Citizens v. City of Santa Ana*, 410 F. Supp. 873, 891–900 (C.D. Cal. 1976) (invalidating a height requirement on the basis of an unjustifiable disparate impact on Mexican American applicants); *Officers for Justice v. Civil Serv. Comm’n*, 395 F. Supp. 378, 380–81 (N.D. Cal. 1975) (invalidating a height requirement because of an unjustifiable adverse impact on Asian and Latino job applicants).

40. According to one recent study, in eleven out of nineteen federal court cases challenging PATs as disparate impact sex discrimination, courts invalidated the procedure at issue. See Wu, *supra* note 4, at 1212 n.82. This success rate compares favorably to other sex discrimination disparate impact claims. *Id.* at 1210–12.

business-necessity defense.⁴¹ Crucially, neither the Supreme Court nor Congress has clarified precisely what is required to show that a hiring device is “job related and consistent with business necessity,”⁴² and the lower courts employ a wide variety of formulations,⁴³ sometimes depending not only on jurisdiction but also on the type of job in question.⁴⁴ Possible interpretations are bracketed at one end by a very deferential standard, sometimes described as a “manifest relationship,” requiring only that the employer could rationally conclude that the test effectively measured attributes that were important to job success.⁴⁵ Bracketing the other end of the spectrum is what has been called a “minimum qualifications” standard,⁴⁶ which requires that the employment test represent an actual floor necessary for successful performance of the job in question.⁴⁷ Some courts have even suggested a bifurcated approach, in which positions that implicate “safety concerns” are scrutinized less closely

41. See, e.g., *Harless v. Duck*, 619 F.2d 611, 615–16 (6th Cir. 1980) (invalidating a test because of a lack of job relatedness); *Blake v. City of Los Angeles*, 595 F.2d 1367, 1375–83 (9th Cir. 1979) (same); *Easterling v. Connecticut*, 783 F. Supp. 2d 323, 335–36 (D. Conn. 2011) (same). But see *Pietras v. Bd. of Fire Comm’rs*, 180 F.3d 468, 474–75 (2d Cir. 1999) (scrutinizing closely whether plaintiff’s evidence satisfied the four-fifths rule).

42. In *Wards Cove Packing Co. v. Atonio*, the Supreme Court departed abruptly from previous interpretations of the business-necessity defense and declared that the “dispositive issue is whether a challenged practice serves, in a significant way, the legitimate employment goals of the employer.” 490 U.S. 642, 659 (1989). In so doing, the Court rejected the contention “that the challenged practice be ‘essential’ or ‘indispensable’ to the employer’s business for it to pass muster.” *Id.* In response to this and other elements of the *Wards Cove* holding, Congress passed the Civil Rights Act of 1991, which rejected the *Wards Cove* formulation and injected the requirement that the practice be not only consistent with business necessity but also “job related.” See 42 U.S.C. § 2000e-2(k)(1)(A)(i) (2012); see also 1 Charles A. Sullivan & Lauren M. Walter, *Employment Discrimination: Law and Practice* § 4.03[C] (4th ed. 2009). However, while the amendment explicitly reinstated the pre-*Wards Cove* law, it failed to elaborate what was meant by job relatedness or business necessity. *Id.*

43. See, e.g., Jerard F. Kehoe & Angela Olson, *Cut Scores and Employment Discrimination Litigation*, in *Employment Discrimination Litigation: Behavioral, Quantitative, and Legal Perspectives* 410, 420–23 (Frank J. Landy ed., 2005) (surveying various standards applied by courts of appeals).

44. For example, in *Spurlock v. United Airlines, Inc.*, the Tenth Circuit wrote:

When a job requires a small amount of skill and training and the consequences of hiring an unqualified applicant are insignificant, the courts should examine closely any pre-employment standard or criteria which discriminate against minorities On the other hand, when the job clearly requires a high degree of skill and the economic and human risks involved in hiring an unqualified applicant are great, the employer bears a correspondingly lighter burden to show that his employment criteria are job-related.

475 F.2d 216, 219 (10th Cir. 1972).

45. See David E. Hollar, *Comment, Physical Ability Tests and Title VII*, 67 U. Chi. L. Rev. 777, 785–87 (2000).

46. Kehoe & Olson, *supra* note 43, at 420.

47. See, e.g., *Lanning v. Se. Pa. Transp. Auth.*, 181 F.3d 478, 489 (3d Cir. 1999).

than those that do not.⁴⁸ The EEOC's Uniform Guidelines take a middle road, requiring that practices be "reasonable and consistent with normal expectations of acceptable proficiency."⁴⁹ The upshot is that courts considering physical-ability requirements apply a panoply of standards to the typically dispositive prong of the disparate impact analysis.

Unlike the disparate impact cases, which challenge the effect of a requirement as discriminatory, cases in the *Bauer* mold challenge the practice of norming itself as a form of discrimination.⁵⁰ Yet the disparate impact cases form the doctrinal backdrop against which employers turn to gender-normed tests like that challenged in *Bauer*, because gender-norming a PAT provides employers a potential solution to the threat of Title VII disparate impact liability by equalizing pass rates. Some of the courts addressing disparate impact claims have affirmatively suggested gender-norming as a permissible alternative to unitary standards. For example, in *Lanning v. Southeastern Pennsylvania Transportation Authority*,⁵¹ the Third Circuit suggested in dicta that the Southeastern Pennsylvania Transportation Authority (SEPTA) "institute a non-discriminatory . . . [aerobic] test that would exclude 80% of men as well as 80% of women through separate aerobic capacity cutoffs for the different sexes."⁵² This, the court asserted, would "help SEPTA achieve its stated goal of increasing aerobic capacity without running afoul of Title VII."⁵³ Indeed, a test that equalizes pass rates does not run afoul of Title VII's disparate impact prohibition.⁵⁴ But whether it runs afoul of Title VII's disparate *treatment* prohibition is another question entirely, to which Part II now turns.

48. Hollar, *supra* note 45, at 787–89; see also *Spurlock*, 475 F.2d at 219.

49. Uniform Guidelines on Employee Selection Procedures, 28 C.F.R. § 50.14 (2017).

50. See *infra* sections IIA, IIC (discussing the test at issue in *Bauer v. Lynch* and the legal theory behind the challenge).

51. 181 F.3d at 478. This case post-dates the adoption of the differential cutoff-score provision by eight years. See Civil Rights Act of 1991, Pub. L. No. 102-166, § 105, 105 Stat. 1071, 1074–75 (codified as amended at 42 U.S.C. § 2000e-2(k) (2012)).

52. *Lanning*, 181 F.3d at 490 n.15.

53. *Id.*

54. This statement requires some qualification in light of the Court's decision in *Connecticut v. Teal*, which rejected the use of a "bottom line" defense to rebut a *prima facie* case of disparate impact discrimination. 457 U.S. 440, 450 (1982). Under *Teal*, if a hiring practice disparately impacts a protected group, it is immaterial that the employer corrects or offsets that disparity elsewhere in the hiring process. See *id.* Some courts have understood the *Teal* bottom-line principle to apply only if the challenged step in the hiring process is a dispositive one—that is, if the component "is an identifiable and dispositive barrier that denies an employment opportunity by preventing an individual from proceeding to the next step in the employment process." *Bradley v. City of Lynn*, 443 F. Supp. 2d 145, 159 (D. Mass. 2006). So a test that excludes the relevant classes at equal rates does not have an impermissible disparate impact so long as the groups clear each dispositive hurdle in the process at equal rates.

II. GENDER-NORMED PHYSICAL-ABILITY TESTS UNDER TITLE VII

Employers wishing to use PATs yet wary of potential disparate impact liability sometimes use gender-normed tests. Unlike unitary standards, which apply a single cutoff across the board, gender-normed tests set different raw cutoff scores for male and female applicants. They thereby circumvent the disparate impact problem described in Part I because male and female applicants pass at roughly equal rates. Most law enforcement organizations in the United States use PATs;⁵⁵ one study estimated that of those, just under a third are gender-normed.⁵⁶

This Part assesses the legality of gender-norming under Title VII, considering the Fourth Circuit's recent decision in *Bauer v. Lynch*,⁵⁷ the first court of appeals case to address the issue. Section II.A lays out the factual background of *Bauer*, illustrating how and why the FBI adopted a gender-normed test. Section II.B provides a brief overview of three threads of Title VII disparate treatment doctrine considered by the *Bauer* court: the Supreme Court's decision in *City of Los Angeles Department of Water & Power v. Manhart*,⁵⁸ the unequal-burdens doctrine, and the statute's prohibition on adjusting test scores. Section II.C then critiques the *Bauer* court's reasoning and concludes that Title VII does not, in fact, permit gender-normed PATs absent a valid business justification, contrary to the Fourth Circuit's conclusion. Finally, section II.D explains why the defense first articulated by the Supreme Court in *Ricci v. Destefano*⁵⁹ cannot excuse an employer's use of a gender-normed test.

A. *Bauer v. Lynch: Factual Background*

Jay J. Bauer was a recent graduate from a Northwestern University master's program when the United States was attacked on September 11, 2001.⁶⁰ Deeply moved by those events, he applied to join the FBI's Special Agent program.⁶¹ The Special Agent program is one of several possible career tracks within the FBI's "Operations and Intelligence" branch; the others include "Intelligence Analysts," "Surveillance," "Forensic Accounting," and "Foreign Languages."⁶² The diverse duties of a Special Agent can include: "work[ing] on matters including terrorism, foreign counterintelligence, cyber-crime, organized crime, white collar crime, public corruption, civil rights violations, financial crime, bribery, bank

55. See Lonsway, *Tearing Down the Wall*, supra note 5, at 257.

56. See id. at 258.

57. 812 F.3d 340 (4th Cir.), cert. denied, 137 S. Ct. 372 (2016).

58. 435 U.S. 702 (1978).

59. 557 U.S. 557 (2009).

60. *Bauer*, 812 F.3d at 344.

61. Id.

62. Career Paths, FBI Jobs, <http://www.fbijobs.gov/career-paths> [<http://perma.cc/ZK38-EWR2>] (last visited Oct. 13, 2017).

robbery, extortion, kidnapping, air piracy, interstate criminal activity, fugitive and drug trafficking matters, and other violations of federal statutes.”⁶³ Admission to the Special Agent program also opens the door to an array of even more specialized professional opportunities. Special Agents may apply to join selective, elite “mission-centric” units, like the Hostage Rescue Team, SWAT, Special Agent Bomb Tech Program, and the Operational Medic Program.⁶⁴

The FBI rejected Bauer’s initial application in 2001, finding his prior work experience lacking.⁶⁵ Bauer returned to school, received a Ph.D., and began work in academia.⁶⁶ In 2008, he reapplied to the FBI; this time, the FBI expressed interest in Bauer’s candidacy, and he began the arduous applicant-screening process.⁶⁷ Bauer excelled during the screening process, which includes several written examinations and oral interviews designed to measure characteristics like communication skill, cognitive capacity, judgment, integrity, and problem-solving ability.⁶⁸ The last step in the application process is the Physical Fitness Test (PFT).⁶⁹

The PFT consists of four events: sit-ups, a 300-meter sprint, push-ups, and a 1.5-mile run.⁷⁰ Based on an internal study, the FBI gender-normed the minimum benchmarks for each of the events to “account for . . . innate physiological differences.”⁷¹ For example, the PFT required male applicants to complete thirty push-ups but required female applicants to complete only fourteen push-ups.⁷² None of these requirements emulated a particular task required of Special Agents.⁷³ Rather, the FBI believed that an applicant’s test results reflected his or her “physical fitness level” and that physical-fitness level was either generally necessary to job performance or, alternatively, a strong indicator that the applicant could complete Special Agent training without injury.⁷⁴

Bauer initially failed the push-up portion of the PFT but passed on his second attempt.⁷⁵ Accordingly, the FBI admitted him to its Special Agent training program at Quantico. Trainees at Quantico must pass each of

63. *Bauer v. Holder*, 25 F. Supp. 3d 842, 849 (E.D. Va. 2014), vacated sub nom. *Bauer*, 812 F.3d 340.

64. Special Agents, FBI Jobs, <http://www.fbijobs.gov/career-paths/special-agents> [<http://perma.cc/9Q9U-CZHB>] [hereinafter Special Agents] (last visited Oct. 13, 2017).

65. *Bauer*, 812 F.3d at 344.

66. *Id.*

67. *Id.*

68. *Id.*; Special Agents, *supra* note 64.

69. *Bauer*, 812 F.3d at 344.

70. *Id.* at 343.

71. *Id.*

72. *Id.* at 344.

73. *Bauer v. Holder*, 25 F. Supp. 3d 842, 863–64 (E.D. Va. 2014), vacated sub nom. *Bauer*, 812 F.3d 340.

74. *Id.*

75. *Bauer*, 812 F.3d at 344.

the twenty-two-week program's four components.⁷⁶ The components are "academics; firearms training; practical applications and skills; and defensive tactics and physical fitness."⁷⁷ Trainees must then re-pass the same PFT administered at the screening stage.⁷⁸ Bauer excelled in training in every area except one: the push-up portion of the PFT.⁷⁹ His peers even selected him president of his class and spokesperson for graduation, yet he was simply unable to complete the thirty push-ups in his five attempts, despite having once passed the test at the screening stage.⁸⁰ On his final try, Bauer fell just one push-up short.⁸¹ As a result, the FBI told Bauer he had three options: resign and leave open the possibility of future employment with the FBI, resign permanently, or be fired.⁸² Bauer filed suit, alleging that the PFT violated Title VII.

B. *The Disparate Treatment Challenge*

Unlike the PAT challenges described in Part I,⁸³ Bauer's challenge did not attack the FBI's test because it disproportionately impacted a protected class. Instead, Bauer argued the test contravened Title VII on two other bases: first, that gender-norming the PFT constituted impermissible disparate treatment on the basis of sex;⁸⁴ and second, that gender-norming the PFT violated Title VII's prohibition on the use of different cutoff scores.⁸⁵

1. *The Manhart Simple Test.* — The basic framework of a disparate treatment challenge under Title VII to a facially discriminatory practice is straightforward.⁸⁶ The statute states, in relevant part, that it is "an unlawful employment practice for an employer . . . to discriminate against any individual . . . because of such individual's . . . sex."⁸⁷ Thus, in a disparate

76. *Id.* at 342.

77. *Id.*

78. *Id.* at 344.

79. *Bauer*, 25 F. Supp. 3d at 848.

80. Bauer even excelled on other parts of the PFT. He scored highly on the 300-meter sprint each time and on the 1.5-mile run. *Id.*

81. *Id.* at 849.

82. *Id.*

83. See *supra* section I.B (describing disparate impact challenges to PATs).

84. *Bauer*, 25 F. Supp. 3d at 854; see also 42 U.S.C. § 2000e-16(a) (2012). Disparate treatment suits against private employers are governed by a different provision of Title VII, 42 U.S.C. § 2000e-2(a), but the legal standards under the two are treated the same. See, e.g., *Brown v. Perry*, 184 F.3d 388, 393 (4th Cir. 1999) (applying case law interpreting § 2000e-2(a) to a § 2000e-16(a) claim).

85. *Bauer*, 25 F. Supp. 3d at 859; see also 42 U.S.C. § 2000e-2(l).

86. By comparison, decisions or policies that are facially neutral but intentionally discriminatory (for example, a subjective determination not to hire a candidate because of her race) violate the disparate treatment prohibition but can be significantly more complex because of the evidentiary issues these cases raise.

87. 42 U.S.C. § 2000e-2(a).

treatment challenge, a plaintiff must first show that a decision or policy was made “because of” sex.⁸⁸ Once the plaintiff has done so, a defendant may still prevail if sex is “a bona fide occupational qualification reasonably necessary to the normal operation of that particular business or enterprise.”⁸⁹ Because facially discriminatory policies almost always of their own force suffice to show a decision “because of sex,”⁹⁰ cases challenging facially discriminatory policies typically turn on the application of the bona fide occupational qualification (BFOQ) defense.⁹¹

The Supreme Court’s decision in *City of Los Angeles Department of Water & Power v. Manhart*⁹² elaborating this general framework is especially relevant to *Bauer*. In *Manhart*, the Court considered whether the pension-contribution policy of the City of Los Angeles, which required higher contributions from women than from men, constituted disparate treatment.⁹³ The city reasoned that women as a class tend to live longer than their male counterparts, and thus, female retirees would on average earn more income from the pension fund.⁹⁴ The Court, accepting as fact that longevity is an empirically proven difference between the sexes, nonetheless rejected the city’s policy as impermissible sex discrimination under Title VII.⁹⁵ It said, in relevant part:

The statute’s focus on the individual is unambiguous. It precludes treatment of individuals as simply components of a racial, religious, sexual, or national class. If height is required for a job, a tall woman may not be refused employment merely because, on the average, women are too short. Even a true generalization about the class is an insufficient reason for disqualifying an individual to whom the generalization does not apply.⁹⁶

Thus, the *Manhart* Court appeared to definitively foreclose reliance on classwide generalizations—even those supported by reliable statistical evidence—as a legitimate basis upon which to distinguish between the sexes, unless that distinction could be justified as a BFOQ. In doing so, it announced a “simple test” for disparate treatment: “whether the evidence shows ‘treatment of a person in a manner which but for that person’s sex would be different.’”⁹⁷

88. *Id.*

89. *Id.* § 2000e-2(e)(1).

90. But see *infra* section II.B.2 (discussing the unequal-burdens doctrine, an exception to this general rule).

91. Sullivan & Walter, *supra* note 42, § 3.01.

92. 435 U.S. 702 (1978).

93. *Id.* at 704.

94. *Id.* at 705.

95. *Id.* at 707–11.

96. *Id.* at 708.

97. *Id.* at 711 (quoting *Developments in the Law, Employment Discrimination and Title VII of the Civil Rights Act of 1964*, 84 Harv. L. Rev. 1109, 1170 (1971)).

2. *An Exception to the Simple Test: The Unequal-Burdens Doctrine.* — Generally, *Manhart's* “simple test” applies in challenges to facially discriminatory policies, and cases turn on the application of the BFOQ.⁹⁸ One exception to this general principle is the “unequal-burdens doctrine.” Since at least the early 1970s, plaintiffs have sought to use Title VII to protect against sex-differentiated appearance and grooming standards.⁹⁹ And from the early years of these challenges until the Supreme Court’s *Price Waterhouse v. Hopkins*¹⁰⁰ decision, lower courts responded almost entirely in one voice: Sex-differentiated appearance standards that apply “equal burdens” are not a form of sex discrimination under Title VII, and employers need not justify them as BFOQs.¹⁰¹ Under this line of reasoning, sex-differentiated appearance standards impose permissible “equal” burdens if the standards are similarly costly, consistent with community norms, and not based on stereotypical notions of differences between the sexes.¹⁰²

The second and third of these requirements are in some tension. Thus, after the Supreme Court’s decision in *Price Waterhouse*, which

98. Sullivan & Walter, *supra* note 42, § 3.01. By comparison, disparate treatment cases that deal with intentional discrimination that may not be facially obvious require application of the burden-shifting framework first articulated in *McDonnell Douglas Corp. v. Green*, 411 U.S. 792, 802 (1973).

99. See, e.g., *Jespersen v. Harrah’s Operating Co.*, 444 F.3d 1104, 1108–10 (9th Cir. 2006) (upholding an elaborate sex-differentiated uniform and grooming policy); *Frank v. United Airlines, Inc.*, 216 F.3d 845, 854–55 (9th Cir. 2000) (invalidating a sex-differentiated weight policy); *Knott v. Mo. Pac. R.R.*, 527 F.2d 1249, 1252 (8th Cir. 1975) (upholding employers’ restriction on hair length for male employees because “slight differences in the appearance requirements for males and females have only a negligible effect on employment opportunities”); see also *infra* note 101 (collecting cases).

100. 490 U.S. 228 (1989).

101. See, e.g., *Fountain v. Safeway Stores, Inc.*, 555 F.2d 753, 756 (9th Cir. 1977) (upholding a requirement that male employees wear ties); *Barker v. Taft Broad. Co.*, 549 F.2d 400, 401 (6th Cir. 1977) (upholding a sex-differentiated hairstyle requirement); *Earwood v. Cont’l Se. Lines, Inc.*, 539 F.2d 1349, 1351 (4th Cir. 1976) (upholding a limit on hair length that applied only to male employees); *Longo v. Carlisle DeCoppet & Co.*, 537 F.2d 685, 685 (2d Cir. 1976) (upholding differential hair requirements for male and female employees); *Knott*, 527 F.2d at 1252 (8th Cir. 1975) (upholding an appearance code with some unitary and some sex-differentiated requirements); *Willingham v. Macon Tel. Publ’g Co.*, 507 F.2d 1084, 1088 (5th Cir. 1975) (emphasizing that Title VII does not prohibit distinctions based on mutable characteristics like hair); *Baker v. Cal. Land Title Co.*, 507 F.2d 895, 898 (9th Cir. 1974) (upholding a differential grooming requirement); *Dodge v. Giant Food, Inc.*, 488 F.2d 1333, 1337 (D.C. Cir. 1973) (upholding a differential hairstyle requirement); see also *Harper v. Blockbuster Ent. Corp.*, 139 F.3d 1385, 1387 (11th Cir. 1998) (upholding differential hair length requirements under *Willingham*, 507 F.2d 1084); *Carroll v. Talman Fed. Sav. & Loan Ass’n of Chi.*, 604 F.2d 1028, 1032 (7th Cir. 1979) (distinguishing the employer’s policy that women wear a “clearly identifiable uniform” while men wear “a variety of normal business attire” from the grooming-standard cases).

102. See *Jespersen*, 444 F.3d at 1109–10 (9th Cir. 2006) (“Grooming standards that appropriately differentiate between the genders are not facially discriminatory.”).

recognized sex stereotyping as actionable under Title VII,¹⁰³ some courts and commentators expressed doubt about the continued vitality of the unequal-burdens doctrine. In particular, commentators argued that *Price Waterhouse's* prohibition on sex stereotyping could not tolerate sex-differentiated grooming standards because these standards typically prescribe conformity with socially constructed, sex-differentiated norms.¹⁰⁴ Some courts, including most prominently the Ninth Circuit in *Jespersen v. Harrah's Operating Co.*, held that *Price Waterhouse* did not affect the unequal-burdens line and that a plaintiff challenging a sex-differentiated grooming standard must show either an impermissible sex stereotype (as distinguished from, apparently, a permissible or de minimis sex stereotype) or unequal burdens in the form of different costs of compliance.¹⁰⁵

In both form and substance, the unequal-burdens doctrine is anomalous. In broad strokes, the structure of Title VII is that a plaintiff may raise the presumption of impermissibility either by showing a formal classification or intentional discrimination (disparate treatment) or by pointing to the discriminatory effects of an otherwise neutral policy (disparate impact).¹⁰⁶ The unequal-burdens doctrine “turn[s] Title VII on its head” by requiring a plaintiff to show *not only* formal or intentional discrimination *but also* discriminatory effects.¹⁰⁷ The *Manhart* Court repudiated this structural inversion, albeit in a different factual context, when it rejected the City of Los Angeles’s argument that its facially discriminatory policy had created no discriminatory effect.¹⁰⁸ Further, it is substantively incompatible with Title VII’s prohibition on sex stereotyping, as embodied in *Price Waterhouse*,¹⁰⁹ because the doctrine perpetuates and fortifies sex stereotypes by allowing sex-differentiated appearance standards only when the distinctions reflect “generally accepted community standards of dress and appearance.”¹¹⁰ Moreover, by measuring “burdens”

103. *Price Waterhouse*, 490 U.S. at 239–42.

104. See, e.g., Joel Wm. Friedman, Gender Nonconformity and the Unfulfilled Promise of *Price Waterhouse v. Hopkins*, 14 Duke J. Gender L. & Pol’y 205, 210–11 (2007) (objecting to sex-differentiated appearance standards as “a type of physical branding or differentiation of female employees that serves to reinforce both the male behavioral norm and the traditionally dominant role enjoyed by men (and the correspondingly subordinate position ascribed to females) in the market place”); Deborah L. Rhode, The Injustice of Appearance, 61 Stan. L. Rev. 1033, 1077 (2009) (“Courts have . . . failed to question the sex stereotypes underlying conventional ‘community standards’ and to demand a reasonable business justification for employers’ restrictions.”).

105. *Jespersen*, 444 F.3d at 1110.

106. See *supra* section I.B (discussing disparate impact cases in the physical-fitness context).

107. Jennifer L. Levi, Misapplying Equality Theories: Dress Codes at Work, 19 Yale J.L. & Feminism 353, 382 (2008).

108. *City of L.A. Dep’t of Water & Power v. Manhart*, 435 U.S. 702, 717–19 (1978).

109. 490 U.S. 228 (1989).

110. *Willingham v. Macon Tel. Publ’g Co.*, 507 F.2d 1084, 1092 (5th Cir. 1975); see also Mary Anne C. Case, Disaggregating Gender from Sex and Sexual Orientation: The

only by monetary costs, courts applying the unequal-burdens doctrine fail to account for the other weighty interests at stake in grooming and appearance cases.¹¹¹

Lower courts justified this apparently atextual exception to Title VII's prohibition on disparate treatment by interpreting "sex" as embracing only *immutable* characteristics.¹¹² This line of reasoning distinguished appearance and grooming standards, such as uniforms, makeup, and hairstyling, as mutable and therefore outside Title VII's purview. While some judges and commentators found this justification less persuasive than others,¹¹³ all courts had at least agreed in one respect: If the unequal-burdens doctrine does apply at all, it applies to only appearance and grooming standards—that is, until the Fourth Circuit's decision in *Bauer*.¹¹⁴

Effeminate Man in the Law and Feminist Jurisprudence, 105 Yale L.J. 1, 48–50 (1995) (arguing that the prohibition on sex stereotyping precludes sex-specific grooming codes); Levi, *supra* note 107, at 356 (describing differential dress codes as the "Title VII blind spot"); Robert Post, Prejudicial Appearances: The Logic of American Antidiscrimination Law, 88 Calif. L. Rev. 1, 30 (2000) ("These cases nicely illustrate how customary gender norms are incorporated into the very meaning and texture of Title VII. . . . Title VII must be understood as marking a frontier between those gender conventions subject to legal transformation and those left untouched or actually reproduced within the law.").

111. See Peter Brandon Bayer, Debunking Unequal Burdens, Trivial Violations, Harmless Stereotypes, and Similar Judicial Myths: The Convergence of Title VII Literalism, Congressional Intent, and Kantian Dignity Theory, 89 St. John's L. Rev. 401, 480–93 (2015) (critiquing unequal-burdens doctrine through the lens of Kantian dignity theory); Noa Ben-Asher, The Two Laws of Sex Stereotyping, 57 B.C. L. Rev. 1187, 1230–31 (2016) ("The problem is that the equal burdens test, however, usually fails to capture the primary harm of grooming and dress codes: harm to personal liberty."); Douglas NeJaime, Marriage Inequality: Same-Sex Relationships, Religious Exemptions, and the Production of Sexual Orientation Discrimination, 100 Calif. L. Rev. 1169, 1208–09 (2012) (critiquing the unequal-burdens doctrine as an example of "the static perspective on discrimination that courts generally use to interpret and apply antidiscrimination law [that] leaves discrimination against conduct-based enactment of identity largely unaddressed").

112. See *Earwood v. Cont'l Se. Lines, Inc.*, 539 F.2d 1349, 1351 (4th Cir. 1976) (upholding an employer's hair-length policy because "[h]air length is not an immutable characteristic for it may be changed at will"); *Willingham*, 507 F.2d at 1092 ("[D]istinctions in employment practices between men and women on the basis of something other than immutable or protected characteristics do not inhibit employment opportunity."); *Dodge v. Giant Food, Inc.*, 488 F.2d 1333, 1336–37 (D.C. Cir. 1973) ("[H]air-length regulations[] are classifications by sex which do not limit employment opportunities by making distinctions based on immutable personal characteristics . . ."); see also Bayer, *supra* note 111, at 414–18 (explaining the origin and rationale of mutability theory).

113. See, e.g., *Barker v. Taft Broad. Co.*, 549 F.2d 400, 404 (6th Cir. 1977) (McCree, J., dissenting) ("[T]he [Supreme] Court [does] not look to the importance, the significance, the mutability, or the fundamental nature of the characteristic that the employer sought to regulate. The Supreme Court limit[s] its inquiry to whether there [is] different treatment of male and female employees.").

114. See *infra* section II.C.

3. *Title VII's Score-Norming Provision.* — The final relevant strain of applicable law concerns a less frequently litigated provision of Title VII¹¹⁵ that prohibits employers from “adjust[ing] the scores of, us[ing] different cutoff scores for, or otherwise alter[ing] the results of, employment related tests on the basis of race, color, religion, sex, or national origin.”¹¹⁶ While the legislators who enacted this provision were principally concerned with race-norming,¹¹⁷ the text proscribes gender-norming as well.¹¹⁸ In fact, after the adoption of the Civil Rights Act of 1991, which added this provision to Title VII, the Cooper Institute, the preeminent developer of PATs in the United States, wrote to law enforcement agencies to state its understanding that the amendment proscribed gender-norming.¹¹⁹ The Department of Justice then repudiated this interpretation.¹²⁰ Courts analyzing gender-normed PATs generally failed to explain at any length why the practice was doctrinally permissible under the post-1991 framework, other than to assert—in seeming conflict with the holding of *Manhart*¹²¹—that distinctions on the basis of “undeniable” physical differences between the sexes were permitted under Title VII.¹²²

C. *Assessing the Fourth Circuit's Reasoning*

The district court in *Bauer* denied the government's motion for summary judgment and found that the different cutoff scores for men

115. Most of the litigation around this provision deals not with whether norming is permitted but rather with whether the challenged practices count as norming at all. See, e.g., *Chi. Firefighters Local 2 v. City of Chi.*, 249 F.3d 649, 656 (7th Cir. 2001) (holding that “banding” doesn't constitute impermissible norming under § 2000e-2(l)).

116. 42 U.S.C. § 2000e-2(l) (2012).

117. See *Hayden v. County of Nassau*, 180 F.3d 42, 53 (2d Cir. 1999) (“The legislative history . . . confirms that it intended to prohibit ‘race norming.’” (citing 137 Cong. Rec. H9529 (daily ed. Nov. 7, 1991); 137 Cong. Rec. S15476 (daily ed. Oct. 30, 1991))). Legislators were to some extent also concerned with gender-norming. See 137 Cong. Rec. 9078 (1991) (statement of Sen. Simpson) (“Chairman Kemp and I have been particularly concerned with the issue of the adjustment of test scores on the basis of race *and sex* and were pleased to see that [the] bill would address this discriminatory practice.” (emphasis added)); *id.* (excerpting criticism by EEOC Vice Chairman R. Gaull Silberman condemning an EEOC recommendation to gender-norm a test as proposing, “[i]n effect, . . . [that] to reduce ‘disparate impact,’ the employer had to hire less qualified, less productive applicants”).

118. 42 U.S.C. § 2000e-2(l).

119. See Letter from Katherine A. Baldwin, Chief, Emp't Litig. Section, Civil Rights Div., U.S. Dep't of Justice, to Roger Reynolds, Assoc. Dir., Cooper Inst. for Aerobics Research (July 29, 1998) (Exhibit 2, Brief in Opposition to Plaintiff's Motion for Class Certification, *Easterling v. Conn. Dep't of Corr.*, 265 F.R.D. 45 (D. Conn. 2010) (No. 3:08-cv-0826 (JCH))) (on file with the *Columbia Law Review*).

120. *Id.*

121. See *supra* section II.B.1 (discussing *Manhart*).

122. See *Powell v. Reno*, No. 96-2743 (NHJ), 1997 U.S. Dist. LEXIS 24169, at *11 (D.D.C. July 24, 1997) (“Physically, the sexes are not ‘similarly situated’; inherent physiological differences exist between them.”).

and women constituted disparate treatment.¹²³ It applied the *Manhart* “simple test”: “[D]iscrimination appears ‘where the evidence shows treatment of a person in a manner which but for that person’s sex would be different.’”¹²⁴ The Fourth Circuit then reversed the district court’s decision. It held that gender-norming the PFT did not constitute impermissible disparate treatment or violate Title VII’s prohibition on different cutoff scores and therefore did not need to be justified as a BFOQ,¹²⁵ so long as the different raw scores represented the same gender-normed fitness level.¹²⁶ It then remanded the case to the district court to determine whether the test did in fact impose equal burdens on each class.¹²⁷

Noting that the appeal involved a “relatively novel issue,” the Fourth Circuit began by setting out what it saw as the “pertinent legal authorities.”¹²⁸ First, it described *Manhart*’s simple test, relied on by the district court.¹²⁹ Then, the court turned to what it described as an alternative to *Manhart*’s simple test: the “no greater burden” test, derived from the unequal-burdens line of grooming and appearance cases.¹³⁰ The court reasoned that the district court was wrong to apply *Manhart*’s simple test because “[m]en and women simply are not physiologically the same for the purposes of physical fitness programs.”¹³¹ This conclusion, the court asserted, was bolstered by the Supreme Court’s dicta in *United States v. Virginia (VMI)*, which recognized in the context of an equal protection challenge that the Virginia Military Institute might adjust fitness standards to accommodate newly admitted female cadets.¹³²

The *Bauer* court’s reliance on the unequal-burdens doctrine is remarkable. The historic limitation of the doctrine to grooming and appearance standards is essential to its “mutability” justification and has been strictly observed.¹³³ Thus, courts considering challenges to sex-differentiated weight requirements have at times expressed hesitation applying the doctrine to a context that is arguably on the outermost limits of “mutability.”¹³⁴ And while courts have occasionally found differential

123. See *Bauer v. Lynch*, 812 F.3d 340, 346–47 (4th Cir.), cert. denied, 137 S. Ct. 372 (2016).

124. *Id.* at 348 (quoting *City of L.A. Dep’t of Water & Power v. Manhart*, 435 U.S. 702, 711 (1978)).

125. *Id.* at 350–51.

126. *Id.*

127. *Id.* at 351–52.

128. *Id.* at 347.

129. *Id.* at 347–48.

130. See *id.* at 349–50.

131. *Id.* at 350.

132. *Id.* (citing *United States v. Virginia (VMI)*, 518 U.S. 515, 550 n.19 (1996)).

133. See *supra* notes 109–114 and accompanying text (discussing courts’ use of the unequal-burdens doctrine before *Bauer*).

134. See, e.g., *Frank v. United Airlines, Inc.*, 216 F.3d 845, 855 (9th Cir. 2000) (noting that a rule “that compels individuals to change or modify their physical structure or

weight requirements acceptable under an equal-burdens logic, they have done so when the requirement reflected an employer's aesthetic preference, not when it was used as a proxy for strength or some other quality.¹³⁵

The *Bauer* court departed dramatically and unceremoniously from this history. The court applied the doctrine, in its own words, because of an innate—read: immutable—difference between the sexes.¹³⁶ But courts have traditionally justified the doctrine's application to grooming and appearance standards precisely because these standards arguably *don't* discriminate on the basis of immutable characteristics.¹³⁷ Unmoored from its traditional limitations and justifications, the doctrine is an atextual and unprincipled enigma. Even if one finds the mutability–immutability justification unsatisfactory, extending the unequal-burdens doctrine to new factual contexts only compounds the problem. It is also at odds with the oft-repeated principle that courts should narrowly interpret exceptions to Title VII liability.¹³⁸

One possible counterargument to this objection would reframe the unequal-burdens doctrine's prior applications into two categories. The first category includes cases that hold that appearance or grooming standards for men and women may be substantively different so long as

composition, as opposed to simply presenting themselves in a neat or acceptable manner” might not “qualif[y] as an appearance standard” but ultimately deciding the case on other grounds).

135. See, e.g., *Gerdomb v. Cont'l Airlines*, 692 F.2d 602, 605–06 (9th Cir. 1982) (holding that an employer's weight policy aimed at employing only “thin, attractive women” was not justifiable unless it imposed equal burdens on men and women).

136. *Bauer*, 812 F.3d at 348. The court's reasoning recalls Professor Katherine Franke's observation:

We have inherited a jurisprudence of sexual equality that seeks to distinguish, as its primary function, inaccurate myths about sexual identity from true—and therefore pre-political—characteristics of sex that are factually significant. As with race, the law uses rules of differentiation to achieve this goal. Yet unlike race, the law reserves a large area for legitimate sex-based regulation—an area bounded by the notion of factually real and legally relevant sexual differences. These rules usually remain unstated, lurking in the background, posing as natural givens, while legal reasoning takes place in the foreground, producing solutions to problems of sex discrimination dependent upon the legitimacy of the essential background assumptions that constitute both the players and the playing field upon which this reasoning process takes place.

Katherine M. Franke, *The Central Mistake of Sex Discrimination Law: The Disaggregation of Sex from Gender*, 144 U. Pa. L. Rev. 1, 29 (1995).

137. See *supra* notes 112–114 and accompanying text (discussing immutability theory).

138. See *infra* note 219 and accompanying text (citing cases establishing that the BFOQ should be interpreted narrowly). Notably, *Bauer* has already been cited for the proposition that “it is an oversimplification to treat [Title VII] as prohibiting any distinction *between* men and women in the workplace. . . . [T]he law prohibits discriminating *against* members of one sex or the other in the workplace.” *Zarda v. Altitude Express, Inc.*, No. 15-3775, 2018 WL 1040820, at *34, *36 (2d Cir. Feb. 26, 2018) (en banc) (Lynch, J. dissenting).

they impose analogous, in the sense of equally costly, burdens. For example, courts have ruled that employers may require female employees to wear makeup if there is a corresponding, but qualitatively different, requirement for male employees—for example, a requirement that they remain clean-shaven.¹³⁹ By comparison, a requirement that female employees wear a uniform with no corresponding requirement whatsoever for male employees would not be acceptable.¹⁴⁰ The second category of cases posits a somewhat different theory: Though the raw scores, cutoffs, or requirements imposed on men and women are quantitatively different, they impose equal, in the sense of qualitatively “the same,” requirements on men and women. Cases dealing with weight requirements—though in fact descended from the mutability justifications—could be reframed as fitting into this distinct category, along with *Bauer*. In this revisionist account, the first class (the substantively-different-but-equally-costly cases) relies on the mutability-immutability distinction, is most compromised by *Price Waterhouse’s* reasoning, and is most vulnerable to criticism on the normative grounds identified by commentators.¹⁴¹ The second class (the formally-different-but-qualitatively-the-same cases, including *Bauer*) is justifiable on the distinct ground that the defendants are in substance imposing the same requirements on men and women. This understanding resolves the tension between the *Bauer* court’s stated justification (innate physiological differences) and the traditional justification for applying the unequal-burdens doctrine (mutable characteristics). It is also consistent with the *Bauer* court’s reliance on *Gerdorn v. Continental Airlines, Inc.*,¹⁴² a case challenging differential weight requirements.

But this understanding raises other questions. The cutoffs in *Bauer* imposed “the same” qualitative burden on male and female applicants only once normed to the applicants’ classes. Thus, the *Bauer* court’s reasoning relied on its determination that “real physiological differences” between the sexes prevent employers from measuring certain qualities (“fitness level” in *Bauer*) without reference to applicants’ sex.¹⁴³ However, *Manhart* seems to foreclose the position that a physiological-differences

139. See, e.g., *Jespersen v. Harrah’s Operating Co.*, 444 F.3d 1104, 1109–10 (9th Cir. 2006) (“While . . . individual [grooming] requirements differ according to gender, none on its face places a greater burden on one gender than the other. Grooming standards that appropriately differentiate between the genders are not facially discriminatory.”).

140. See *Carroll v. Talman Fed. Sav. & Loan Ass’n of Chi.*, 604 F.2d 1028, 1032 (7th Cir. 1979) (invalidating a uniform requirement for female employees on this ground).

141. See *supra* notes 108–111 and accompanying text (explaining the tension between *Price Waterhouse* and the unequal-burdens doctrine).

142. 692 F.2d 602 (9th Cir. 1982).

143. See *supra* section II.B.1 (reviewing the *Manhart* decision).

rationale can negate a prima facie case.¹⁴⁴ The *Bauer* court distinguished *Manhart*, or at least justified setting aside its simple test in favor of the unequal-burdens analysis, based on the observation that physiological differences exist between the sexes.¹⁴⁵ Yet, the key question in *Manhart* was whether a “physiological” difference between the sexes—longevity—could justify adjustments in conditions of employment.¹⁴⁶ The *Manhart* Court plainly rejected that argument, reasserting the focus on the individual envisioned by Title VII.¹⁴⁷ The *Bauer* court failed to explain why this pronouncement did not squarely address and dispose of the defendant’s theory on the import of statistical “physiological” discrepancies.¹⁴⁸

Perhaps one could reconceptualize the *Manhart* decision as rejecting the employer’s reliance on physiological differences because of the *particular* generalization and practice at issue. In other words, *Manhart* could be understood as saying that the nexus between the practice—in essence, making women go home with smaller paychecks at the end of the day—and the physiological difference—longevity—is too remote. But this reading seems flatly contrary to the language of the opinion. More fundamentally, distinguishing *Bauer* on these grounds only begs the question. The core issue in *Bauer* was whether an employer who uses a gender-normed PAT in hiring should be required to show some nexus

144. *Bauer*, for his part, argued that the Supreme Court’s decision in *UAW v. Johnson Controls, Inc.*, 499 U.S. 187 (1991), rejected “physiological differences” as a legitimate basis for differentiation between the sexes. Brief for Plaintiff-Appellee at 18, *Bauer v. Lynch*, 812 F.3d 340 (4th Cir. 2016) (No. 14-2323), 2015 WL 2147710. This response is not wholly satisfactory, though, as the Court in *Johnson Controls* found that the sexes actually *were* similarly situated, physiologically speaking, with respect to the relevant employment opportunity, because lead exposure had deleterious effects on the reproductive systems of both men and women. *Johnson*, 499 U.S. at 198. .

145. *Bauer*, 812 F.3d at 350.

146. See *supra* notes 93–94 and accompanying text (discussing the facts of *Manhart*).

147. See *supra* notes 95–97 and accompanying text (describing *Manhart*’s core holding).

148. This may also raise concerns about courts co-opting this rationale to exclude or discriminate against transgender individuals based on “real differences.” Just a few months after the *Bauer* decision, a district court in the Fourth Circuit cited to *Bauer*’s reasoning to justify excluding transgender people from the public bathroom that corresponds to their gender identity:

[A]s recently as January 2016, the Fourth Circuit . . . conclud[ed] that physiological differences justified treating men and women differently in some contexts. In *Bauer*, . . . [t]he Fourth Circuit found that . . . the FBI could distinguish between men and women on the basis of physiology In light of the foregoing, it appears that the privacy interests that justify the State’s provision of sex-segregated bathrooms, showers, and other similar facilities arise from physiological differences between men and women, rather than differences in gender identity.

Carcaño v. McCrory, 203 F. Supp. 3d 615, 643 (M.D.N.C. 2016).

between the test and the job at issue to justify the practice.¹⁴⁹ Under the *Bauer* framework, an employer using a normed test with no disparate impact would not have to justify the practice either as job-related and consistent with business necessity or as a BFOQ. Thus, the nexus between the physiological difference and the consequence could be as remote as in *Manhart*.

The *Bauer* court's proposed framework is also susceptible to abuse. The court accepted as "undeniable" the fact that there exists some abstract concept of "fitness" that (a) is related to trainees' performance and (b) can be measured only on a normed basis.¹⁵⁰ But the court did not cite any evidence presented by either party on these points, and there is reason to doubt those assumptions. For example, one study of female and male army trainees indicates that fitness level is in fact an important indicator of injury risk, the harm the FBI sought to avoid.¹⁵¹ Yet the study defined fitness level by raw, non-normed scores and found gender to be an insignificant predictor of injury risk when controlling for fitness level.¹⁵² Thus, a woman with a given 1.5-mile run time was as likely to be injured as a man *with the same run time*, not with the same gender-normed performance.¹⁵³ It also indicated, consistent with several other studies, that although women entered the training program in significantly worse physical shape than men, female trainees made much bigger fitness gains during the training program.¹⁵⁴ This study compromises two of the *Bauer* court's factual assumptions: that "fitness level," as measured relative to one's gender class, is a relevant indicator of injury risk, and that differences in fitness between male and female trainees are necessarily or entirely attributable to innate physiological differences. These may seem like small quibbles with the facts of the *Bauer* case, and surely a single study is not dispositive of the issue; yet, this finding illustrates how courts applying the unequal-burdens doctrine to "physiological differences" cases might rely on erroneous assumptions about which differences are "real" or relevant. An employer seeking to defend a practice under the *Bauer* framework need only cook up some abstract construct and assert that the construct can be measured only on a gender-normed basis to escape Title VII's requirement of a business justification.

The *Bauer* court also argued that the same unequal-burdens reasoning that justified its finding that there was no disparate treatment also

149. See *infra* section III.A (noting that the effect of the *Bauer* approach is to negate employers' need to show a business justification).

150. *Bauer*, 812 F.3d at 348 (citing *Powell v. Reno*, No. 96-2743, 1997 U.S. Dist. LEXIS 24169, at *9-10 (D.D.C. July 24, 1997)).

151. Nicole S. Bell et al., *High Injury Rates Among Female Army Trainees: A Function of Gender?*, *Am. J. Preventative Med.*, Apr. 2000, at 141, 145.

152. *Id.* at 142-43.

153. *Id.*

154. *Id.* at 145.

justified its finding that the policy did not contravene Title VII's cutoff-score provision.¹⁵⁵ Yet the cutoff-score provision would be meaningless if it required only that individuals receive the same score relative to their respective class. Indeed, Congress adopted the test-cutoff provision specifically due to its concern with norming.¹⁵⁶ Instead of addressing this question head-on, the court seemed to imply that its observation about "physiological differences" justified this conclusion as well.

All in all, the *Bauer* court's reasoning fails to explain how Title VII permits gender-normed PATs without any valid business justification. The unequal-burdens test is an atextual, normatively problematic branch of Title VII doctrine, based on the shaky distinction between "mutable" and "immutable" characteristics. This already-tenuous justification cannot explain the extension of the doctrine to embrace a distinction based on "innate" physiological differences. Further, the court's reliance on physiological differences seems directly contrary to the Supreme Court's holding in *Manhart*. Finally, the facts of *Bauer* demonstrate how the doctrine is susceptible to abuse and how it encourages courts to rely on armchair empiricism about which physiological differences are real or relevant.

D. *The Ricci Defense*

As something of an aside, one final strain of Title VII doctrine deserves brief attention. The district court in *Bauer* suggested that the defense first articulated in *Ricci v. DeStefano*¹⁵⁷ might alternatively provide an out for employers using gender-normed tests. Because the FBI had not raised a *Ricci* defense, the district court did not examine its applicability at length, and the Fourth Circuit did not consider it at all.¹⁵⁸ Yet it is doubtful that *Ricci* can or should apply here.

In *Ricci*, white firefighters challenged the decision of the city of New Haven, Connecticut, to throw out the results of two promotion exams after their administration.¹⁵⁹ The city decided to do so because the exams had had an unexpectedly extreme disparate impact on African American applicants.¹⁶⁰ The Court's decision proceeded in several steps: First, it assumed without discussion that the decision to throw out the tests constituted disparate treatment.¹⁶¹ Second, it held that such a decision could be justified only if there was a "strong basis in evidence" of disparate

155. *Bauer*, 812 F.3d at 350.

156. For legislative history supporting this claim, see *supra* note 117.

157. 557 U.S. 557 (2009).

158. *Bauer v. Holder*, 25 F. Supp. 3d 842, 860 (E.D. Va. 2014), vacated sub nom. *Bauer*, 812 F.3d 340; *Bauer*, 812 F.3d at 346 n.7.

159. *Ricci*, 557 U.S. at 563–74.

160. *Id.*

161. See *id.* at 579.

impact liability but for the decision.¹⁶² Finally, it applied each of the disparate impact prongs (prima facie case, job relatedness–business necessity, and alternative, less discriminatory practice) to the hypothetical case-within-a-case to assess whether the defendant would have been liable had it not thrown out the exams.¹⁶³ Upon finding there was no strong basis in evidence for the potential disparate impact liability, the Court invalidated the decision to throw out the test results.¹⁶⁴

There are at least two reasons why *Ricci* should not or would not apply to a case like *Bauer*.¹⁶⁵ First, in *Ricci*, the original test had been administered and the results received; New Haven’s attempt to avoid disparate impact liability was based on these concrete results. The Court was therefore able to assess on the record before it whether the employer would have been liable but for its decision to scrap the test results. In *Bauer*, by contrast, it’s not clear what the relevant counterfactual for the case-within-the-case would be. Would a court simply assume that the employer would otherwise adopt a unitary standard as stringent as the higher of the two cutoffs? The lower of the two? Or would it look to the test the employer used prior to the gender-normed approach, no matter how far back in history it must go or how related that former test is to the challenged one? The point is that *Ricci* works only if there is a factual baseline on which to judge the hypothetical disparate impact liability.

162. *Id.* at 583–84.

163. *Id.* at 585–92.

164. *Id.* at 593.

165. In an influential analysis of *Ricci*, Professor Richard Primus argues that the case can be interpreted in one of three ways. See Richard Primus, *The Future of Disparate Impact*, 108 *Mich. L. Rev.* 1341, 1344 (2010). The first interpretation is the “general reading.” *Id.* On this view, classification of any kind by any governmental actor—including courts—is a violation of statutory and constitutional antidiscrimination mandates. *Id.* at 1363. Thus, in Primus’s words, courts may not “classify members of the workforce by race in order to adjudicate disparate impact claims” and “equal protection requires the law to be thoroughly colorblind.” *Id.* (Though Primus’s analysis uses the facts of the *Ricci* case and the language of race discrimination, his statutory and equal protection arguments would apply with equal force to sex discrimination.) In other words, in this extraordinarily broad interpretation, the mere act of adjudicating a disparate impact claim could itself constitute disparate treatment. As the continued adjudication of disparate impact claims demonstrates, this view has not (at least yet) prevailed. The second interpretation is the “institutional reading.” *Id.* at 1364. The institutional reading “holds that courts may order race-conscious remedies for disparate impact problems, but public employers may not.” *Id.* The final interpretation is the “visible-victims reading,” which draws on an antibalkanization rationale to hold that “race-conscious measures that visibly burden specific innocent parties”—in *Ricci*, high scorers on the original test—are qualitatively different than measures “intended to improve the position of disadvantaged groups but whose costs are more diffuse.” *Id.* at 1371; see also Reva B. Siegel, *From Colorblindness to Antibalkanization: An Emerging Ground of Decision in Race Equality Cases*, 120 *Yale L.J.* 1278, 1345–47 (2011) (identifying Primus’s visible-victims reading as a form of antibalkanization). The difficult question of *Ricci*’s precise contours and consequences need not be resolved, though, to demonstrate that at a minimum, it cannot apply to a case like *Bauer*; indeed, none of Primus’s three theories would preclude that conclusion.

Second, a more fundamental issue is whether *Ricci* can be stretched so far as to permit the adoption of a test that itself is a form of disparate treatment. In *Ricci*, the problematic practice was the (one-time) decision to throw out the original test after administration, *not the new test adopted after that decision*.¹⁶⁶ But in *Bauer* and cases like it, the problematic practice is the test itself.¹⁶⁷ It would radically expand *Ricci*'s reach to apply it to excuse this kind of ongoing disparate treatment. Taken to these extremes, *Ricci* could be read to excuse any decision or policy adopted "because of" a protected characteristic, so long as there is some imaginable counterfactual that would give rise to disparate impact liability.¹⁶⁸

III. A SOLUTION TO THE GENDER-NORMED PHYSICAL-ABILITY TEST PROBLEM: A RETURN TO TITLE VII FIRST PRINCIPLES

The analysis in Part II suggests that, contrary to the *Bauer* court's reasoning, Title VII does not permit the use of gender-normed PATs absent a valid business justification. This Part considers whether this is a desirable state of affairs. Section III.A argues that the *Bauer* court's approach is harmful to women's equality in the workplace and that requiring employers to put forth a business justification for the use of gender-normed PATs helps root out entrenched stereotypes about the primacy of masculinity in traditionally male job sectors. Section III.B then suggests that applying the traditional framework for disparate treatment claims—in other words, requiring a demonstration that the desired characteristic is a bona fide occupational qualification—best addresses this concern. Yet, the application of the disparate treatment framework to gender-norming must be paired with a demanding business-necessity standard in the corresponding disparate impact challenges to unitary PATs. A rule that requires a similarly rigid business justification under either a disparate impact theory or a disparate treatment theory would incentivize employers to tailor their physical hiring practices more closely to the actual demands of the job, thereby discouraging arbitrary

166. *Ricci*, 557 U.S. at 592.

167. See *Bauer*, 812 F.3d at 860.

168. One interesting and perhaps more relevant coda to *Ricci*: After the case concluded, an African American firefighter in New Haven brought the underlying disparate impact suit that the *Ricci* Court had hypothesized. *Briscoe v. City of New Haven*, 967 F. Supp. 2d 563, 564–65 (D. Conn. 2013). During litigation, the city argued that even if the plaintiff prevailed on the disparate impact claim, the court would be limited in fashioning a remedy by 42 U.S.C. § 2000e-2(l), the Title VII cutoff-score provision. *Id.* at 592–93. The district court rejected the argument and instead held that the cutoff-score provision restricted employers ("respondents" in the statute's terms) but did not cabin courts' discretion to prescribe norming as a remedy. *Id.* at 592. If broadly adopted, this holding would permit courts to order gender-norming as a remedy to disparate impact claims even though employers may not be able to do so under the framework outlined in section III.B. This holding also lends support to Primus's institutional reading of *Ricci*. See Primus, *supra* note 165, at 1364.

practices that either promote harmful stereotypes or have discriminatory effects.

A. *An Antisubordination Critique of the Bauer Approach*

Scholars have framed debates over the proper approach to anti-discrimination law as a conflict between the anticlassification and antisubordination traditions.¹⁶⁹ The core claims of the anticlassification approach are that distinctions on the basis of a protected characteristic are virtually never permissible and that facially neutral practices are virtually always permissible, so long as they are not mere pretexts to invidious discrimination.¹⁷⁰ It is therefore sometimes described as the “colorblindness” principle when applied to race discrimination¹⁷¹ and is associated with the conservative wing of the Supreme Court.¹⁷² In contrast, the antisubordination approach holds that antidiscrimination law should aim to combat the historic and systematic subordination of certain classes and that distinctions on the basis of race, sex, or other protected characteristics are not objectionable if they seek that end.¹⁷³ It is associated with the liberal wing of the Supreme Court.¹⁷⁴ Recently, Professor Reva Siegel has proposed replacing this traditional dyadic model with a triadic model, recognizing a third position: the antibalkanization

169. See generally Jack M. Balkin & Reva B. Siegel, *The American Civil Rights Tradition: Anticlassification or Antisubordination?*, 58 U. Miami L. Rev. 9 (2003).

170. For foundational work in the anticlassification tradition, see generally Paul Brest, *The Supreme Court, 1975 Term—Foreword: In Defense of the Antidiscrimination Principle*, 90 Harv. L. Rev. 1 (1976).

171. See Ian F. Haney López, “A Nation of Minorities”: Race, Ethnicity, and Reactionary Colorblindness, 59 Stan. L. Rev. 985, 987–88 (2007) (“[T]he Supreme Court in the last three decades has moved ever closer to a full embrace of an anticlassification or colorblind conception of the Equal Protection Clause.”).

172. See Siegel, *supra* note 165, at 1282–83. The anticlassification tradition holds considerably less sway in the academy but has been influential in Supreme Court antidiscrimination jurisprudence. See Samuel R. Bagenstos, *The Structural Turn and the Limits of Antidiscrimination Law*, 94 Calif. L. Rev. 1, 40–41 (2006) (“In the courts, the antidiscrimination principle dominates.”).

173. See Owen M. Fiss, *Groups and the Equal Protection Clause*, 5 Phil. & Pub. Aff. 107, 157–60 (1976) (“The concern should be with those laws or practices that particularly hurt a disadvantaged group. . . . [W]hat is critical . . . is that the state law or practice aggravates (or perpetuates?) the subordinate position of a specially disadvantaged group.”); see also Ruth Colker, *Anti-Subordination Above All: Sex, Race, and Equal Protection*, 61 N.Y.U. L. Rev. 1003, 1007–10 (1986) [hereinafter Colker, *Anti-Subordination*] (“From an anti-subordination perspective, both facially differentiating and facially neutral policies are invidious only if they perpetuate racial or sexual hierarchy.”); Cass R. Sunstein, *The Anticaste Principle*, 92 Mich. L. Rev. 2410, 2428–33 (1994) (“Instead of asking ‘Are blacks or women similarly situated to whites or men, and if so have they been treated differently?’ we should ask ‘Does the law or practice in question contribute to the maintenance of a second-class citizenship, or lower-case status, for blacks or women?’”).

174. See Siegel, *supra* note 165, at 1282–83.

principle.¹⁷⁵ Unlike the anticlassification perspective, this approach recognizes a moral and legal difference between those measures adopted to promote equality and those adopted to reinforce inequality; yet, unlike the antisubordination perspective, it gives credence to concerns about social cohesion and prefers state action that is neutrally structured to mitigate social provocation.¹⁷⁶ Professor Siegel divines this approach from the opinions of “swing” Justices, including Justices Powell, O’Connor, and Kennedy.¹⁷⁷

Title VII challenges advanced by favored-group members like Jay Bauer (a man bringing a sex discrimination claim) pose distinctive problems for antisubordination theories.¹⁷⁸ These so-called “reverse discrimination” suits—which comprise a significant portion, if not the majority, of recent challenges to law enforcement hiring¹⁷⁹—seem to pit interests in substantive equality, the chief antisubordination concern, against formal equal treatment, the chief anticlassification concern. Indeed, the use of different cutoffs is flatly unacceptable from an anticlassification perspective. The legislators who championed Title VII’s cutoff-score provision spoke in essentially anticlassificationist terms,¹⁸⁰ and the *Bauer* decision drew anticlassificationist criticism from right-leaning media sources because of the perceived hypocrisy of treating sex discrimination claims by men differently than those by women.¹⁸¹ Thus, the *Bauer* dilemma might, at first blush, seem like a neat microcosm of the debate between the antisubordination and anticlassification approaches. The following analysis attempts to complicate that understanding.

In a simplistic sense, gender-norming can open doors for women. Compared against the baseline of a unitary standard employing the higher of the two possible cutoffs—for example, thirty push-ups in *Bauer*—

175. *Id.*, *passim*.

176. *Id.* at 1300–03.

177. See *id.* at 1303 (“Justice Kennedy stakes out a position in the tradition of Justices Powell and O’Connor that is responsive to the tug of each vision, while refusing cleanly to adopt either.”).

178. See generally Robert K. Fullinwider, *The Reverse Discrimination Controversy: A Moral and Legal Analysis* (1980); Sullivan, *supra* note 22, at 1506 (describing the growing acceptance of “reverse discrimination” claims as a reflection of “the increasingly reflexive colorblindness in our law”).

179. See Liyah Kaprice Brown, *Officer or Overseer?: Why Police Desegregation Fails as an Adequate Solution to Racist, Oppressive, and Violent Policing in Black Communities*, 29 N.Y.U. Rev. L. & Soc. Change 757, 773–77 (2005) (discussing trends in reverse discrimination suits against police departments).

180. See *supra* note 117 (quoting senatorial debate on the provision).

181. See, e.g., Ed Whelan, *Transgressive Progressives*, Nat’l Rev. Online: Bench Memos (Jan. 21, 2016), <http://www.nationalreview.com/bench-memos/430136/title-vii-sex-discrimination-transgender> [<http://perma.cc/7Y4U-HZKL>] (“As two recent federal court rulings indicate, progressives will give an unnaturally stingy reading of Title VII when men (or men *qua* men, I suppose I must say) allege discrimination and an adventurously expansive reading when members of their favored constituencies do so.”).

gender-norming may permit more women to access certain employment opportunities. Nonetheless, gender-norming PATs can meaningfully, though indirectly, contribute to the subordination of women in the workplace, particularly in traditionally male-dominated occupations. The practice insulates from judicial scrutiny¹⁸² employment selection devices that privilege masculine physicality, even though those devices may have little to do with the job at issue. In the law enforcement context, these practices arbitrarily buttress the stereotype that women are innately less capable of successfully filling roles that implicate public safety.¹⁸³ This stereotype negatively and concretely impacts women's application to, retention in, and promotion from these roles.¹⁸⁴

Under the *Bauer* framework, employers may use selection devices that have no relation whatsoever to the job at issue, so long as the employers have successfully normed the relevant cutoff scores. To illustrate, an employer seeking to fill a role that requires no physical strength at all—say, an accounting job—could require male and female applicants to pass a gender-normed push-up test without contravening Title VII. On the other hand, if an employer administered a unitary push-up test that had a statistically disparate impact, the employer almost certainly could not justify it as job related and consistent with business necessity.¹⁸⁵ And even if the employer could somehow overcome that barrier, the plaintiff could still prevail by showing the existence of a less discriminatory alternative.

The accountant hypothetical seems absurd because push-ups are so obviously unrelated to accounting, but this hypothetical, exaggerated as it may be, is not quite as dissimilar to law enforcement as it appears. In the words of Professor Mary Anne Case, “The job of police officer is one whose history of being gendered masculine is virtually unsurpassed.”¹⁸⁶

182. While this Note emphasizes judicial scrutiny, there are, of course, other mechanisms by which Title VII is enforced and effectuated, most importantly agency enforcement by the EEOC and voluntary compliance by employers.

183. See Case, *supra* note 110, at 81–94 (condemning the harmful and discriminatory “use of gendered traits in the construction of a job and in the selection, training, and evaluation of those who perform the job”); Ruth Colker, Rank-Order Physical Abilities Selection Devices for Traditionally Male Occupations as Gender-Based Employment Discrimination, 19 U.C. Davis L. Rev. 761, 796–801 (1986) [hereinafter Colker, Physical Abilities Selection] (making a similar argument in the context of unitary standards and disparate impact claims). See also *infra* note 192 and accompanying text (citing research demonstrating the success of women in law enforcement roles).

184. See *infra* notes 205–207 and accompanying text (citing research demonstrating the deleterious effect of stereotyping on women's success in law enforcement).

185. See *supra* notes 42–49 and accompanying text (describing the job-relatedness–business-necessity defense).

186. See Case, *supra* note 110, at 85. Case's analysis emphasizes two kinds of antidiscrimination cases against police agencies: first, disparate impact challenges to PATs, like those discussed in section I.B, see *id.*; and second, disparate treatment cases brought by individual women “challenging the perception, rooted in stereotypes, that they and

Yet the relationship between strength and speed and successful law enforcement is unclear at best, and the literature on the relation of physical selection devices to performance in law enforcement roles is mixed.¹⁸⁷ For many, the idea of policing conjures images of dramatic physical work, but the bulk of police work doesn't involve intense physical tasks like foot pursuits—indeed, modern policing standards often recommend against them¹⁸⁸—and doesn't (or perhaps shouldn't) routinely involve violent confrontations.¹⁸⁹ What it does involve is crisis

their kind are unsuited to police work." *Id.* (citing *Thorne v. City of El Segundo*, 726 F.2d 459, 464–65 (9th Cir. 1983)).

187. Compare Richard D. Arvey et al., Development of Physical Ability Tests for Police Officers: A Construct Validation Approach, 77 *J. Applied Psychol.* 996, 1008 (1992) (arguing there is "fairly convincing evidence for the construct validity of a set of physical ability test events" to select entry-level police officers), with *id.* (acknowledging that the study was at least partially limited by the risk that supervisors' performance ratings "simply respond[ed] to stereotypes based on body size and body fat . . . [or] gender."), and Amie M. Schuck, Female Representation in Law Enforcement: The Influence of Screening, Unions, Incentives, Community Policing, CALEA, and Size, 17 *Police Q.* 54, 70 (2014) [hereinafter Schuck, Female Representation] ("Physical fitness requirements are controversial because there is little empirical evidence that these tests are (a) reflective of the most common tasks that officers are expected to perform, (b) predictive of performance when coping with hostile or noncompliant citizens, and (c) associated with fewer negative organizational outcomes . . .").

188. See Robert J. Kaminski et al., Police Foot Pursuits: Report on Findings from a National Survey on Policies, Practices, and Training 1–4 (2012), http://researchgate.net/publication/280805424_Police_foot_pursuits_Report_on_findings_from_a_national_survey_on_policies_practices_and_training (on file with the *Columbia Law Review*) ("[C]oncerns about foot-pursuit related injuries and fatalities, in part, led to . . . [the] release [of] a model foot-pursuit policy in 2003 The policy also recommends that lone officers not try to overtake a fleeing suspect to make an arrest.").

189. Indeed, renewed outrage over police brutality has reinvigorated public scrutiny of the relationship between the gender composition of police agencies and episodes of police violence. See, e.g., Shaun King, To Combat Police Brutality, Hire More Female Cops—Studies Show They're Better at Keeping Their Cool, *N.Y. Daily News* (Aug. 3, 2016), <http://www.nydailynews.com/news/national/king-combat-police-brutality-hire-female-cops-article-1.2737314> (on file with the *Columbia Law Review*) (noting the vast gender disparity in police forces and arguing that a more gender-balanced workforce would decrease the incidence of police brutality); Amy Stewart, Opinion, Female Police Officers Save Lives, *N.Y. Times* (July 26, 2016), <http://www.nytimes.com/2016/07/26/opinion/female-police-officers-save-lives.html> (on file with the *Columbia Law Review*) (same). Other recent public controversies have also sparked fresh interest in the composition of federal law enforcement agencies. See, e.g., Nick Baumann, Maybe the FBI's Love for Trump Has Something to Do with How Extremely White and Male It Is, *Huffington Post* (Nov. 4, 2016), http://www.huffingtonpost.com/entry/fbi-trump-white-male_us_581cc321e4b0aac62483f6e4 [<http://perma.cc/N3SF-QN64>] (noting that the FBI is disproportionately white and male and that it has actually gotten more racially unrepresentative in recent years); Adam Goldman, Where Are Women in F.B.I.'s Top Ranks?, *N.Y. Times* (Oct. 22, 2016), http://www.nytimes.com/2016/10/23/us/fbi-women.html?_r=0 (on file with the *Columbia Law Review*) (citing data showing that women hold only twelve percent of the FBI's senior agent positions, a decline from twenty percent in 2013).

de-escalation; communication with victims, suspects, and communities; clerical work; and patrol.¹⁹⁰

Even when police officers are called on to address violent crime, strength and speed requirements are still suspect predictors of performance. Some advocates of a more gender-balanced police workforce emphasize that most violent crime to which officers respond is male-on-female domestic violence and that women as a class are better at responding effectively to these situations.¹⁹¹ A related body of scholarship also suggests that, on average, female police officers are more adept at avoiding violent confrontations in the first instance.¹⁹² These accounts admittedly speak in an essentialist register that should engender skepticism. But the claim here is not that women make better police officers or different “kinds” of police officers. Rather, the claim is a more modest one: that courts should not simply assume that masculine-coded traits like strength and speed are necessary to effective performance of a job—even one as seemingly familiar as law enforcement—without demanding some evidence that that is so.

Anecdotal evidence also illustrates this point. As discussed in section I.B, PATs frequently fail the job-relatedness–business-necessity test in disparate impact challenges, showing that it is hardly farfetched to believe that employers are using invalid physical selection devices.¹⁹³ *Bauer* too is illustrative. In *Bauer*, the FBI justified its PFT on two bases: First, it argued that the test was important to strong on-the-job performance.¹⁹⁴ The district court rejected this justification, noting that the absence of any physical test for incumbent Special Agents belied the argument that the PFT was necessary to success in that role.¹⁹⁵ In the alternative, the FBI argued the PFT was a screening device intended to

190. See Anastasia Prokos & Irene Padavic, ‘There Oughtta Be a Law Against Bitches’: Masculinity Lessons in Police Academy Training, 9 *Gender Work & Org.* 439, 441–43 (2002) (“Male police officers have drawn on images of a ‘masculine cop’ to enhance their sense of masculinity and to resist women’s growing presence The reality of police work, however, involves much tedium and paperwork and relatively little crime fighting or violence.”).

191. See Kim Lonsway et al., *The Nat’l Ctr. for Women & Policing, Men, Women, and Police Excessive Force: A Tale of Two Genders* 9 (2002), http://womenandpolicing.com/PDF/2002_Excessive_Force.pdf [<http://perma.cc/2UER-T5CC>].

192. See Amie M. Schuck, *Gender Differences in Policing: Testing Hypotheses from the Performance and Disruption Perspectives*, 9 *Feminist Criminology* 160, 161 (2014) [hereinafter Schuck, *Gender Differences*] (collecting studies showing that female police officers are less likely to receive complaints about excessive use of force, are better at avoiding violent confrontations, are better at de-escalating confrontations, and are better at communicating with victims).

193. See *supra* section I.B (discussing this case law).

194. *Bauer v. Holder*, 25 F. Supp. 3d 842, 863 (E.D. Va. 2014), vacated sub nom. *Bauer v. Lynch*, 812 F.3d 340 (4th Cir.), cert. denied, 137 S. Ct. 372 (2016).

195. *Id.* at 864.

filter out applicants prone to injury during the training program.¹⁹⁶ Evidence showing that the FBI initially developed the PFT in response to high injury rates partially corroborated this claim.¹⁹⁷ Yet, Bauer passed the *pre*-training PFT.¹⁹⁸ It was the *post*-training test that he was unable to pass.¹⁹⁹ In other words, he had already avoided training injury by the time he failed, rendering the FBI's justification nonsensical.²⁰⁰

PATs' often suspect relation to job performance is especially significant because these tests frequently demand physical qualities like strength and speed that code as masculine, as opposed to physical qualities like flexibility and endurance that code as feminine or gender neutral.²⁰¹ This poses a significant impediment to employment equality.²⁰² First, it can deter female job seekers from applying to these jobs.²⁰³ Second, the use of these tests emphasizes qualities that women are perceived to possess in lesser amounts than men, generating and perpetuating the view that women are inherently less qualified to serve in these roles.²⁰⁴ In the law enforcement context, a significant body of

196. *Id.*

197. *Id.*

198. *Id.*

199. *Id.*

200. *Id.*

201. See Colker, *Physical Abilities Selection*, *supra* note 183, at 793–96 (describing the decision of the New York City Fire Department to ignore professional advice to adopt standards that measured physical traits that code as feminine, like flexibility, and instead use strength and speed tests).

202. See Janet Chan et al., *Doing and Undoing Gender in Policing*, 14 *Theoretical Criminology* 425, 426 (2010) (“[T]raditional policing takes for granted the crime-fighting and coercive nature of police work and equates policing with physicality. This in turn leads to the assumption that policing is naturally a man’s job Being female therefore has the potential to carry negative symbolic capital in the field of policing.” (citations omitted)).

203. See Danielle Gaucher et al., *Evidence that Gendered Wording in Job Advertisements Exists and Sustains Gender Inequality*, 101 *J. Personality & Soc. Psychol.* 109, 119–20 (2011); Schuck, *Female Representation*, *supra* note 187, at 69–70 (finding that “organizational policies and practices appear to have a greater impact on the representation of women in law enforcement than community factors . . . outside law enforcement” and “reaffirm[ing] the negative impact of physical fitness requirements on female representation in law enforcement”). Significantly, the Supreme Court acknowledged the deterrent effect of discriminatory hiring devices in *Dothard v. Rawlinson*, 433 U.S. 321, 330 (1977). The defendants in that case contended that the disparate impact of the height and weight requirements was not as dramatic as the plaintiffs alleged because the relevant comparator was not the general population of the United States but rather the pool of actual applicants to the Alabama correctional officer positions. *Id.* The Court rejected that argument, recognizing the distorting effect of the selection device: “The application process itself might not adequately reflect the actual potential applicant pool, since otherwise qualified people might be discouraged from applying because of a self-recognized inability to meet the very standards challenged as being discriminatory.” *Id.*

204. The perceived inadequacy of female police officers is a well-documented phenomenon with myriad real-world consequences for women and for the communities

literature shows that these stereotypes undermine policewomen's credibility among peers²⁰⁵ and the communities they serve,²⁰⁶ professional advancement, and work satisfaction.²⁰⁷

When applied to gender-normed tests, this latter critique may partially sound in antibalkanization. The antibalkanization theory suggests that preferential treatment exacerbates interclass resentment and erodes social cohesion. This, in turn, undermines the advancement of protected classes, whose achievements are tainted with the odor of paternalistic preference.²⁰⁸ Thus, on an antibalkanization account, gender-norming might stunt policewomen's success by arousing resentment among peers and superiors. But the critique also has an important antisubordination angle. On this account, the trouble with gender-normed tests is that they shield employers' potentially arbitrary use of hiring practices that reinforce harmful and unnecessary stereotypes about women's ability to perform in the workplace. Further, gender-normed tests arguably amplify the problem, by emphasizing in one breath the importance of masculinity to success, while reminding in the next breath that women can't possibly measure up.

With that said, one need not think that police work doesn't require much strength or speed to believe the antisubordination critique of *Bauer*. The point is simply that the costs described above are justifiable

the police seek to protect. See Chan et al., *supra* note 202, at 425–26; Prokos & Padavic, *supra* note 190, at 439; Schuck, *Gender Differences*, *supra* note 192, at 166.

205. See Robin N. Haarr, *Patterns of Interaction in Police Patrol Bureau: Race and Gender Barriers to Integration*, 14 *Just. Q.* 53, 71 (1997) (describing male officers' belief that female officers received lighter workloads and less dangerous assignments); Merry Morash & Robin N. Haarr, *Doing, Redoing, and Undoing Gender: Variation in Gender Identities of Women Working as Police Officers*, 7 *Feminist Criminology* 3, 16 (2012) (recounting evidence of a negative stereotype among male police officers that their female peers are inferior due to physical inadequacy); Prokos & Padavic, *supra* note 190, at 453–54 (describing the disrespect accorded female instructors by male recruits); Schuck, *Gender Differences*, *supra* note 192, at 161 (“Research indicates that some male officers doubt that women can adequately perform the tasks associated with the occupation, often questioning their physical and emotional capabilities.”).

206. Morash & Haarr, *supra* note 205, at 4 (“Existing research does reveal persistent tendencies for the public and for some police officers to equate effective policing with crime fighting by a person with ‘masculine’ capacities for ‘aggression, violence, danger, risk taking, and courageousness.’” (quoting Cortney A. Franklin, *Male Peer Support and the Police Culture: Understanding the Resistance and Opposition of Women in Policing*, 16 *Women & Crim. Just.*, no. 3, 2005, at 1, 6)).

207. See Jenny Veldman et al., *Women (Do Not) Belong Here: Gender-Work Identity Conflict Among Female Police Officers*, 8 *Frontiers in Psychology* 1, 6 (2017) (finding gender isolation among female members of police teams led to “a stronger perception that their team members see their gender as conflicting with their work identity” which ultimately led to “more burn-out symptoms, less extra role behavior, lower job satisfaction, lower work motivation, and lower perceived performance”).

208. See generally Siegel, *supra* note 165, at 1334–36 (“[H]owever majority group grievement differs from minority group grievement, it nevertheless can stimulate racial resentments that erode social cohesion.”).

only if the job at issue *actually requires* the physical qualities for which the employer tests. Put differently, one can, consistent with this view, believe (1) that employers should adopt only those physical selection devices that accurately predict job performance and (2) that gender-normed law enforcement PATs are likely to pass this test. The question is merely whether employers should have to show the relation at all.

Finally, to be clear, non-job-related hiring procedures that emphasize traditionally masculine or feminine qualities do not alone violate Title VII absent either impermissible disparate treatment or disparate impact. Title VII requires business justification not in the first instance but rather in response to a *prima facie* case of discrimination.²⁰⁹ Thus, if gender-norming PATs is not itself a form of disparate treatment, the antisubordination critique would not of its own force transform those tests into a form of sex discrimination under Title VII. However, if gender-norming is a form of disparate treatment under current Title VII doctrine, as argued in Part II, the critique above suggests it should not be understood as a straightforward case of anticlassification or antibalkanization triumphing over antisubordination. Or, put another way, the critique suggests that judges inclined to read Title VII through an antisubordination lens need not strain present doctrine to accommodate the practice. While unitary hiring standards that impose a disparate impact on women perpetuate the gender hierarchy by exclusion, job-unrelated gender-normed PATs perpetuate the gender hierarchy by arbitrarily privileging masculinity while evading judicial review. *Bauer* may then be the rare “reverse discrimination” case in which anticlassification, antibalkanization, and antisubordination point to the same result.

B. *Applying Title VII to Gender-Normed Physical-Ability Tests*

In proposing a solution to the puzzle of gender-normed PATs, this section proceeds from the following premises: First, unitary fitness standards that have a disparate impact on a protected class are a form of discrimination, unless they are job related and consistent with business necessity.²¹⁰ Second, arbitrary hiring practices that reinforce stereotypes about women’s inadequacy impose a distinct harm, even if they do not have a disparate impact on protected classes.²¹¹ Third, a *Bauer*-like approach to gender-norming insulates these practices from judicial review, thus creating a problem from an antisubordination perspective (in addition to the more obvious anticlassification and antibalkanization critiques).²¹² And fourth, courts should adopt neither the unequal-

209. See *supra* sections I.A, II.B (explaining the basic disparate impact and disparate treatment frameworks).

210. See *supra* Part I.

211. See *supra* section III.A.

212. See *supra* section III.A.

burdens doctrine nor any other account based on “real physiological differences” to accommodate this practice.²¹³

Based on these premises, an ideal regime would permit employers to adopt PATs—normed or unitary—if, but only if, success on those tests were truly critical to the performance of the job at issue. Requiring employers to justify gender-normed and unitary PATs does not put the employer in an impossible damned-if-you-do, damned-if-you-don’t bind, since a third alternative—eliminating or decreasing the physical requirements—always remains open. Existing Title VII doctrine provides just such a solution. Specifically, physical hiring tests with an impermissible disparate impact should be assessed under a demanding business-necessity standard that would require a showing that the test reflects the actual requirements of the job. This is the approach already taken by some courts of appeals.²¹⁴ This strict standard for disparate impact challenges should be paired with the typical disparate treatment framework in normed cases; in other words, employers would need to show that their gender-normed PATs were BFOQs.

The application of the BFOQ defense to a gender-normed PAT presents two wrinkles. Traditionally, the question the BFOQ defense poses is: Is sex itself a bona fide occupational qualification “reasonably necessary to the normal operation of that particular business or enterprise[?]”²¹⁵ The question posed in the gender-norming context would be: Is the desired gender-normed quality a BFOQ? There is no theoretical barrier to applying the test in that way, but there arguably is a textual one; the statute, in describing the BFOQ defense, says it is not unlawful for an “employer to hire and employ employees . . . on the basis of his . . . sex . . . in those certain instances where . . . sex . . . is a bona fide occupational qualification”²¹⁶ In contrast to other portions of the statute, it says nothing explicit to excuse an employer “classifying” employees on the basis of sex,²¹⁷ nor does it say anything about employment actions other than hiring or refusing to hire. But it would be absurd to read this provision to excuse employers from firing on the basis of sex and to excuse employment agencies from classifying on the basis of sex, but not to excuse employers from classifying on the basis of sex.²¹⁸

213. See *supra* section II.C.

214. See *supra* note 46 and accompanying text (describing the “minimum qualifications” standard).

215. *UAW v. Johnson Controls, Inc.*, 499 U.S. 187, 212 (1991) (internal quotation marks omitted) (quoting 42 U.S.C. § 2000e-2(e) (2012)); see also *Dothard v. Rawlinson*, 433 U.S. 321, 333–35 (1977).

216. 42 U.S.C. § 2000e-2(e).

217. The same provision does excuse employment agencies from “classifying” on the basis of sex if sex is a BFOQ. *Id.*

218. The author is unaware of any case apart from *Bauer* in which a court has addressed whether (or even assumed that) a BFOQ may excuse a violation of Title VII’s

The trickier question is when, if ever, gender-normed physical fitness would be a BFOQ. The BFOQ defense is exceedingly strict by design.²¹⁹ A requirement must “relate to the ‘essence,’ or to the ‘central mission of the employer’s business’” to be a BFOQ.²²⁰ For example, the safety of third parties who are neither customers nor essential to the business cannot support a BFOQ;²²¹ similarly, neither the additional cost of employing one sex²²² nor customer preference counts as a BFOQ.²²³ That’s not to say that a gender-normed PAT logically never could be a BFOQ—just that it’s difficult to imagine when it would be. But the stringency of the test does not undermine the reasoning supporting its application.²²⁴

At most, the strict construction of the BFOQ defense in disparate treatment cases underscores the need for a demanding business-necessity standard in corresponding disparate impact cases. Requiring employers who gender-norm to justify the tests as BFOQs but then giving what would amount to a free pass²²⁵ to employers who use unitary standards would in fact exacerbate the problem outlined in section III.A. It would likely have the additional effect of decreasing access for women by allowing employers to set arbitrarily high, exclusionary unitary standards but prohibiting them from lowering those standards when applied to women. Thus, courts taking the approach advocated here and rejecting the *Bauer* court’s analysis of gender-normed tests must bring a similar skepticism to unitary standards and require an actual, demanding showing of business necessity and job relatedness. In sum, traditional Title VII disparate treatment doctrine provides the proper framework under which

cutoff-score provision. See 42 U.S.C. § 2000e-2(*l*) (2012). This dearth of case law might be attributable to the conceptual difficulties arising from the application of the BFOQ defense to norming. See *infra* notes 219–224 and accompanying text. There may be a simpler explanation, though: the vast majority of cases brought under that provision are race discrimination claims, rather than sex discrimination claims, and there is no BFOQ defense to race discrimination. As a practical matter, though, the applicability of the BFOQ to the cutoff-score provision is largely immaterial because it is unlikely that an employer would be able to show that the normed score was a BFOQ. See *infra* notes 219–224 and accompanying text.

219. *Johnson Controls*, 499 U.S. at 201 (“The BFOQ defense is written narrowly, and this Court has read it narrowly.”); see also, e.g., *Teamsters Local Union No. 117 v. Wash. Dep’t of Corr.*, 789 F.3d 979, 987 (9th Cir. 2015); *Everson v. Mich. Dep’t of Corr.*, 391 F.3d 737, 747 (6th Cir. 2004).

220. *Johnson Controls*, 499 U.S. at 203 (first quoting *Dothard*, 433 U.S. at 333; then quoting *W. Airlines, Inc. v. Criswell*, 472 U.S. 400, 413 (1985)).

221. *Id.* (rejecting the safety of an employee’s unborn fetuses as a basis for a BFOQ).

222. *Id.*

223. *Diaz v. Pan Am. World Airways, Inc.*, 442 F.2d 385, 389 (5th Cir. 1971).

224. See *supra* section III.A.

225. See *supra* note 45 and accompanying text (describing the very lenient manifest relationship standard used by some courts of appeals).

to analyze gender-norming, but only if courts require a business justification in corresponding disparate impact cases as well.

CONCLUSION

The Fourth Circuit's decision in *Bauer v. Lynch* stretches the unequal-burdens doctrine beyond its principled limits. In doing so, it creates a significant and anomalous exception to Title VII and shields from judicial scrutiny arbitrary hiring practices that reinforce harmful stereotypes about women's inadequacy in public safety roles. As this Note suggests, the better reading of Title VII's text and Supreme Court jurisprudence is that gender-normed PATs are not permitted absent a BFOQ. This doctrinal conclusion is normatively defensible not only from an antibalkanization or anticlassification perspective but also from an antistatutory perspective, and courts considering gender-normed physical-ability tests in the future should reject the *Bauer* court's approach. But this solution will serve its intended ends only if courts apply a stringent business-necessity standard in corresponding challenges to unitary physical hiring criteria, an approach that some but not all lower courts already take. By asking employers that insist on using discriminatory physical-ability tests in any form to present business justifications, courts can best promote Title VII's antistatutory principle.

