

---

## ESSAY

---

---

### WHAT TAYLOR SWIFT AND BEYONCÉ TEACH US ABOUT SEX AND CAUSES

---

ROBIN DEMBROFF, ISSA KOHLER-HAUSMANN & ELISE SUGARMAN<sup>†</sup>

#### INTRODUCTION

One of the most anticipated decisions of this term will be the three consolidated cases pending in front of the Supreme Court, *Bostock v. Clayton County*, *Altitude Express v. Zarda*, and *R.G. & G.R. Harris Funeral Homes v. EEOC*, which collectively present the question of whether Title VII's prohibition on discrimination "because of sex" includes discrimination against gay, lesbian, and transgender employees. On October 8, 2019, Prof. Pamela Karlan's oral argument for *Zarda* and *Bostock* centered around one basic hypothetical:

[W]hen you tell two employees who come in, both of whom tell you they married their partner Bill last weekend, when you fire the male employee who married Bill and you give the female employee who married Bill a couple of days off so she can celebrate the joyous event, that's discrimination because of sex.<sup>1</sup>

---

<sup>†</sup> Robin Dembroff, Assistant Professor Philosophy, Yale University. Issa Kohler-Hausmann, Professor of Law & Sociology, Yale Law School. Elise Sugarman, J.D. candidate, Yale Law School; PhD candidate, Philosophy, Stanford University. The authors wish to thank Jonathan Schaffer and Mark Schroeder for helpful conversation and feedback during the development of this paper, and Martha Fitzgerald, Lily Hu, Lisa Hogan, Esteban Morin, Amanda Shanor, Katie White, and Gideon Yaffe for their help in thinking through a related amicus brief submitted by two of the authors.

<sup>1</sup> Transcript of Oral Argument at 7-8, *Bostock v. Clayton County*, No. 17-1618 (U.S. argued Oct. 8, 2019).

The hypothetical, which stunned the Justices into a momentary silence, was meant to demonstrate satisfaction of a test canonically enunciated in *Manhart*.<sup>2</sup> As Justice Kagan mentioned to opposing counsel,

*Manhart* gave us a very simple test, and *Manhart* said, what you do when you look to see whether there is discrimination under Title VII is, you say, would the same thing have happened to you if you were of a different sex? . . . We have insisted on this extremely simple test.<sup>3</sup>

The *Manhart* test, which asks “whether the evidence shows treatment of a person in a manner which but for that person’s sex would be different,”<sup>4</sup> is doctrinally referred to as a “but-for” causal test for sex discrimination.<sup>5</sup> Such counterfactual thought experiments are inherently indeterminate as to whether the act or practice is discrimination, and the precise form of the counterfactual will always be contested. Defendants argue, for example, that Prof. Karlan’s counterfactual involves changing more than *just* the employee’s sex when we ask if the treatment would have been different. “[S]ex and sexual orientation are different traits,”<sup>6</sup> defendants claim, so the latter should be fixed when imagining changes to the former. For this reason, “[t]he correct comparison is between a female employee in a same-sex relationship and a male employee in a same-sex relationship.”<sup>7</sup> Supplying another example, they argue that

[t]he proper analysis [for] a neutral policy, such as use [of] the showering facility that corresponds to your biological sex, [for] the man who uses the women’s shower, the . . . comparator is not a woman who uses the woman’s

---

<sup>2</sup> In *City of Los Angeles, Department of Water and Power v. Manhart*, the Los Angeles Department of Water and Power conditioned costs for their monthly pension plan on sex because female employees generally had longer life expectancy than their male counterparts. 435 U.S. 702, 711 (1978). The Court held that the policy ran afoul of Title VII, explaining that “[s]uch a practice does not pass the simple test of whether the evidence shows treatment of a person in a manner which but for that person’s sex would be different.” *Id.* (internal quotation omitted).

<sup>3</sup> Oral Argument at 35:11, *Bostock*, No. 17-1618 (U.S. argued Oct. 8, 2019), [https://www.supremecourt.gov/oral\\_arguments/audio/2019/17-1618](https://www.supremecourt.gov/oral_arguments/audio/2019/17-1618) [<https://perma.cc/Y89G-2KNR>]. Note that the official transcript incorrectly attributes this statement to Justice Sotomayor. Transcript of Oral Argument, *supra* note 1, at 41-42.

<sup>4</sup> *Manhart*, 435 U.S. at 711 (internal quotation omitted).

<sup>5</sup> We are not claiming either party consistently adheres to one formulation of the test in their arguments. See, for example, a slightly different version of Prof. Karlan’s thought experiment in the opening lines of oral argument:

When a[n] employer fires a male employee for dating men but does not fire female employees who date men, he violates Title VII. The employer has, in the words of Section 703(a), discriminated against the man because he treats that man worse than women who want to do the same thing.

Transcript of Oral Argument, *supra* note 1, at 4.

<sup>6</sup> *Id.* at 61; see also Brief for Respondent at 7, *Bostock*, No. 17-1618 (U.S. Aug. 16, 2019) (arguing that sexual orientation and sex are separate and distinct concepts).

<sup>7</sup> Brief for the United States as Amicus Curiae at 10, *Bostock*, No. 17-1618 (U.S. Aug. 23, 2019).

shower. It's a woman who uses the men's shower, because otherwise . . . you're not looking at similarly situated people.<sup>8</sup>

Although differences in the formulation of the counterfactual yield different normative implications, the but-for causal test assumes that sex discrimination can be revealed by counterfactually “toggling” the sex features of an individual plaintiff.<sup>9</sup> Both plaintiffs and defendants assume, in short, that so long as we have the right idea of what individual features we must change in the counterfactual, we will have the capacity to imagine a “similarly situated” person who differs from the plaintiff *only* with respect to those individual features.

These thought experiments sometimes yield what, by our lights, is the right answer to the question presented in these cases—that firing someone for being gay or gender nonconforming *does* constitute discrimination because of sex.<sup>10</sup> However, the logic of but-for causal tests is misguided in a way that obfuscates the real legal and normative questions animating antidiscrimination cases. Although both parties' argumentative strategies are constrained by pre-existing doctrinal formulations that rely upon this type of causal reasoning, this essay shows that but-for causal tests confuse more than clarify a legal inquiry into whether or not something is discriminatory. They ought to be replaced with a more coherent approach. We argue in Part 1 (Negative Argument) that it is a mistake to think that we can answer questions of whether or not something is an instance of discrimination by asking about *individual-level causation* (i.e., the but-for causal test), which centers on inherent traits or attributes of individual plaintiffs. We propose in Part 2 (Positive Argument) that *social explanation* is more appropriate for identifying instances of discrimination, as it centers on the social generalizations, stereotypes, norms, and expectations (hereon “social meanings”) attached to sex categories.

Beyoncé and Taylor Swift illustrate the difference between individual-level causation and social explanation in two separate songs, “If I Were a Boy” and “The Man.”<sup>11</sup> Covering similar themes, Beyoncé sings: “If I were a boy . . . I'd kick it with who I wanted, and I'd never get confronted for it . . . .”<sup>12</sup> In “The Man,” Taylor Swift sings:

---

<sup>8</sup> Transcript of Oral Argument, *supra* note 1, at 49.

<sup>9</sup> In this essay, we use the term “sex features” to refer exclusively to physical features of individuals that are associated with male or female reproductive roles.

<sup>10</sup> In the interest of full disclosure, two of the authors here submitted an amicus brief in support of the employees that were fired on behalf of over seventy philosophers arguing this. See Brief of Philosophy Professors as Amici Curiae in Support of the Employees, *Bostock*, No. 17-1618 (U.S. July 3, 2019).

<sup>11</sup> BEYONCÉ, *If I Were a Boy*, on I AM . . . SASHA FIERCE (Columbia Records 2008); TAYLOR SWIFT, *The Man*, on LOVER (Republic Records 2019).

<sup>12</sup> BEYONCÉ, *supra* note 11.

[I]f I was a man . . . they wouldn't shake their heads and question how much of this I deserve . . . What I was wearing, [and] if I was rude could all be separated from my good ideas and power moves . . . [We] would toast to me . . . let the players play, I'd be just like Leo in Saint-Tropez.<sup>13</sup>

Beyoncé and Swift are not claiming that having male sex features per se causes an individual to get away with “kick[ing] it with who [one] wants” or to be perceived as a deserving player. Rather, they are telling us that “kick[ing] it with who [one] wants” and not getting confronted for it, being presumed to deserve one's success, or being considered a player instead of promiscuous is explained by the social categories of sex. That is, the locus of explanation for why the “male” versions of Beyoncé and Swift are perceived differently does not lie in the individual-level features of sex considered apart from the social world, but in social-level roles and expectations associated with those features. Failing to recognize this distinction, one might mistakenly conclude from Beyoncé's and Swift's lyrics that it is something about the individual person, not society, that explains the discriminatory outcome. But as Beyoncé's and Swift's bodies of work explore, the *social meanings* of sex categories—the generalizations, stereotypes, norms and assumptions associated with these categories—really do the explanatory work.<sup>14</sup>

An allegation of discrimination under Title VII demands a social explanation. However, while a social explanation can tell us when social meanings of a category explain an outcome, it alone does not tell us if this outcome is *wrongful* or, more specifically, *discriminatory*. To answer that question, we would need a normative theory that tells us which employment practices are permissible, and which ought to be prohibited, given that sex categories impose different expectations, norms, roles, etc. onto people based on their sex classification. Disagreement over normative theories of this sort, we think, is the real debate animating the parties' competing formulations of counterfactuals. The plaintiff-employees hold that a person ought to be protected from adverse employment action notwithstanding the fact that they violate prevailing stereotypes and expectations regarding (real or presumed) sexual attraction and dress/presentation. The defendant-employers hold that they ought to be free to take adverse employment action against an employee who violates prevailing stereotypes and expectations regarding sexual attractions and dress/presentation. No value-neutral inquiry into causality or explanation can settle this disagreement.

---

<sup>13</sup> TAYLOR SWIFT, *supra* note 11.

<sup>14</sup> See, e.g., BEYONCÉ, *LEMONADE* (Columbia Records 2016); TAYLOR SWIFT, *You Need to Calm Down*, on *LOVER* (Republic Records 2019).

## I. NEGATIVE ARGUMENT: AGAINST INDIVIDUAL-LEVEL CAUSAL EXPLANATION

When judges ask what would have occurred if a female plaintiff had been male, or a male plaintiff had been female, they engage in a test for individual-level causal explanation. By that we mean a test that seeks to identify whether an act or practice is discriminatory by asking if it is causally explained by the individual plaintiff's sex features.

Individual-level causation is a concept largely taken from torts. In torts, the law asks whether an isolated event or action caused a particular outcome, or what is sometimes referred to as “event” or “token” causation.<sup>15</sup> An accident case where *D* drove his car into *V* requires us to identify if the event of *D*'s driving as he did was a cause of a distinct later event, namely *V*'s injury. Here, the counterfactual thought experiment is useful, because what we want to know is: if *D* had not thus driven the car, would *V* not have been injured?<sup>16</sup> The candidates for causes are what we imagine toggling in the counterfactual world. This is, more generally, how counterfactual causal tests work: in order to test whether *A* caused *B*, we imagine a counterfactual world in which *A* did not occur and see whether *B* still occurred.<sup>17</sup>

When the law applies this framework to the question of discrimination, it attempts to do something similar, but by toggling individuals' features rather than events. The law asks: in a counterfactual world where those features have been altered, would the outcome have been different?<sup>18</sup> This framework is inappropriate to the domain of discrimination, which concerns questions about social categories, and not isolated features of individuals. Instances of sex

---

<sup>15</sup> See JESSICA COLLINS, NED HALL & L.A. PAUL, CAUSATION AND COUNTERFACTUALS 254 (Jessica Collins, Ned Hall & L.A. Paul eds., 2004) (“[W]hen we can say that the fact that *e* occurred depends on the fact that *c* occurred—then we can go ahead and call this a kind of event-causation.”).

<sup>16</sup> There is significant debate about which competing concept of causation best captures the causal relations we should be concerned with in tort cases, and this debate points out that these but-for tests are often implicitly relying on a very specific contrastive. For excellent discussions on this debate, see, for example, Jonathan Schaffer, *Contrastive Causation in the Law*, 16 LEGAL THEORY 259 (2010), and Note, *Rethinking Actual Causation in Tort Law*, 130 HARV. L. REV. 2163 (2017).

<sup>17</sup> And of course, even when we can easily identify but-for causation in a tort case, that does not settle the normative question of liability because both the plaintiff and defendant were usually but-for causes of the accident!

<sup>18</sup> See, for example, *Comcast Corp. v. National Ass'n of African American-Owned Media*, 140 S. Ct. 1009 (2020), in which the Supreme Court held that the but-for test is an appropriate causal requirement throughout the life of a lawsuit brought under 42 U.S.C. § 1981 (2018). The Court reasoned that “[t]he guarantee that each person is entitled to the ‘same right . . . as is enjoyed by white citizens [to make and enforce contracts]’ directs our attention to the counterfactual—what would have happened if the plaintiff had been white?” *Id.* at 1015. In that case, the Court refers to the but-for test as “textbook tort law”—an “ancient and simple” test that “supplies the ‘default’ or ‘background’ rule against which Congress is normally presumed to have legislated when creating its own new causes of action.” *Id.* at 1014 (citation omitted).

discrimination, for example, are not explained by individuals' sex features. Rather, they are explained by the social meanings and expectations imposed onto people on the basis of these features. A person with male-coded sex features who is fired for having a feminine presentation is not fired because of their sex features *per se*, but because they are viewed as defying the social meanings of those features. A man who is fired for having a male partner is not fired because of his sex features *per se*, but because he is viewed as defying social meanings of those features (e.g., norms of attraction only to women). And so on.

Although prior work has recognized that these counterfactual thought experiments are indeterminate as to the normative question of discrimination, we can explain why that is the case.<sup>19</sup> Sex features cannot be isolated from their social meanings. A counterfactual scenario wherein an individual has different sex features sneaks in these social meanings and, with them, the substantive significance of the individual's other attributes or behaviors relative to prevailing sex-specific norms and expectations. Anytime we imagine changing *just* the "trait" of sex (we assume they imagine the "trait" of sex consists in a person's reproductive sex features, e.g., changing a penis to a vagina), but holding constant the complained-of trait (e.g., "wearing a dress" or "presumed sexual attractions to males"), we are *necessarily* changing the meaning of the trait in light of the sex features (now, for example, the person is gender *conforming* with respect to sexuality and dress).<sup>20</sup> These examples illustrate a more general truth: how these counterfactuals are formulated is a conceptual judgment call with normative implications. This is precisely *why* such thought experiments are inherently indeterminate, and why the parties and Justices spent so much time fighting about the details of the relevant counterfactual.<sup>21</sup> We cannot change the traits that make someone a member of a given sex category without also

---

<sup>19</sup> See, e.g., David A. Strauss, *Discriminatory Intent and the Taming of Brown*, 56 U. CHI. L. REV. 935 (1989). And as Justice Breyer has pointed out,

It is one thing to require a typical tort plaintiff to show "but-for" causation. In that context, reasonably objective scientific or commonsense theories of physical causation make the concept of "but-for" causation comparatively easy to understand and relatively easy to apply. But it is an entirely different matter to determine a "but-for" relation when we consider, not physical forces, but the mind-related characterizations that constitute motive.

Gross v. FBL Fin. Servs., Inc. 557 U.S. 167, 190 (2009) (Breyer, J., dissenting).

<sup>20</sup> Said another way, when the employers insist that "sex and sexual orientation are different traits," they fail to understand that an employee's sex is precisely what determines whether or not the employee's sexual orientation is considered normative or non-normative of social expectations. See *supra* note 6 and accompanying text.

<sup>21</sup> This also explains why we see a startling variety in how but-for counterfactuals are trotted out—in what features are and aren't taken to travel with the category swap. For example, conservatives don't like very narrow counterfactuals in the context of sex discrimination because they sweep in conduct they think is *not* discriminatory such as sex-segregated bathrooms or locker rooms; but they do like them in the context of race discrimination because they sweep in conduct they think *is* discriminatory like affirmative action.

changing the social meanings and expectations imposed on that person. This, in turn, changes the substantive significance of their other attributes.

This point illuminates the central flaw with the defendants' arguments in *Zarda*, *Bostock*, and *R.G. & G.R. Harris Funeral Homes*. Defendants argue that the appropriate, "similarly situated" comparator to use in the but-for test is someone of the opposite sex who violates the gender stereotypes of *that* sex.<sup>22</sup> The amicus for the Trump Department of Justice argues that a counterfactual in which we imagine a female employee attracted to females were instead a male employee attracted to females would "change[] *both* the sex (from female to male) *and* sexual orientation (from gay to straight)" of that employee. "A proper comparison," they maintain, "would change the sex while holding the sexual orientation constant. Only a relative difference in treatment in *that* scenario would constitute sex discrimination."<sup>23</sup> And defendant R.G. & G.R. Harris Funeral Homes argues:

It is only when a court is "scrupulous about holding *everything* constant except the plaintiff's sex" that the comparator analysis can "do its job of *ruling in* sex discrimination as the actual reason for the employer's decision." A woman who identifies as a woman not only has a different sex (female) than Stephens, but also a different transgender status (nontransgender). Such a comparison cannot show sex discrimination.<sup>24</sup>

Clearly, none of these examples are "scrupulous[ly] . . . holding everything constant except the plaintiff's sex." To the contrary, in order to hold constant the complained of conduct—here, sexuality or identity-based gender nonconformity—they counterfactually change the plaintiff's sex features *as well as* their preferred sexual partners or gender identity. The but-for causal inquiry is not an independent test of discrimination; it is an expression of one's normative priors about what is discriminatory/wrongful vis-a-vis sex categories.<sup>25</sup> The defendants insist on the counterfactual specification that

---

<sup>22</sup> See, e.g., Transcript of Oral Argument, *supra* note 1, at 49. The defendants provide the example that

[t]o isolate sex, *Zarda* (a man attracted to the same sex) must be compared to a lesbian woman (a woman attracted to the same sex). Because employers that base decisions on employees' sexual attraction would treat both *Zarda* and the lesbian comparator the same way, the comparator analysis reveals no sex discrimination. Neither sex is favored over the other.

Brief for Petitioners at 9, *Altitude Express, Inc. v. Zarda*, No. 17-1623 (U.S. Aug. 16, 2019).

<sup>23</sup> Brief for the United States as Amicus Curiae, *supra* note 7, at 19.

<sup>24</sup> Brief for the Petitioner at 26-27, *R.G. & G.R. Harris Funeral Homes, Inc. v. EEOC*, No. 18-107 (U.S. Aug. 16, 2019) (quoting *Hively v. Ivy Tech Cmty. Coll. of Ind.*, 853 F.3d 339, 366 (7th Cir. 2017) (Sykes, J., dissenting)).

<sup>25</sup> This explains the inconsistent approach to how the counterfactuals are trotted out. For example, conservatives propose counterfactuals that entail changing sex features in addition to sex-normative

maintains the plaintiffs' noncompliant relationships to gender norms because they believe that, for example, firing a masculine-presenting woman is not discriminatory on the grounds that a similarly gender nonconforming man (i.e., a feminine-presenting man) also would have been fired.<sup>26</sup>

The defendants attempt to obscure these normative priors with a sleight of hand that changes how abstractly the counterfactual adjustments are described. Rather than hold fixed individual features (e.g., "sexual attraction to males," "female gender identity"), they hold fixed relational features (e.g., "sexual orientation," "transgender") that in turn hold constant the plaintiff's gender nonconformity.<sup>27</sup> Such stacking the deck by manipulating the counterfactual just proves our point, which is that *individual-level* traits cannot explain why a gay or transgender employee was fired. The defendants recognize that *social-level* expectations about dress, presentation, sexuality, etc. are responsible; it is why they insist on holding plaintiffs' gender nonconformity fixed, rather than their sexual attraction to males or female gender identity.

## II. POSITIVE ARGUMENT: IN FAVOR OF SOCIAL EXPLANATION

The tangle of counterfactual thought experiments is not mysterious at all once we recognize that the statuses that Title VII forbids from being the basis of discrimination do not consist in merely individual features (i.e., ones that mark a person as an instance of a kind of sex, race, religion, and so on). Rather, they consist in memberships in social categories—categories brimming with often nefarious social meanings. It is, in fact, the *purpose* of antidiscrimination law to revise these nefarious meanings, and to protect individuals from discrimination on the basis of these meanings. For example, when the Supreme Court recognized in *Price Waterhouse v. Hopkins* that firing a woman for being "overly aggressive" could be an instance of sex discrimination, it is

---

conduct in the context of sex discrimination, because more narrow counterfactuals (that just change sex features) would sweep in conduct they think is *not* discriminatory such as sex-segregated bathrooms or locker rooms. However, they propose more narrow counterfactuals that entail changing only markers of racial membership (e.g. skin color) in the context of race discrimination, because doing so sweeps in conduct they think *is* discriminatory, such as affirmative action.

<sup>26</sup> See *id.* at 36-37 (describing how an employer's "belief about transgender status is not sex-specific" because it "applies equally to all men and women"); see also Transcript of Oral Argument, *supra* note 1, at 53 ("So if you treat all . . . gay men and women exactly the same regardless of their sex, you're not discriminating against them because of their sex."); Brief for Petitioners, *supra* note 22, at 9.

<sup>27</sup> For example, the defendants stated that

[i]t is also wrong to say that the proper comparator for Stephens is a biological female who wants to dress as a female. Stephens is a transgendered biological male, so the proper comparator is a transgendered biological female. Changing the comparator's sex and transgender status fails to demonstrate that Harris treats male employees differently than similarly situated female employees. It is a shell game—not a tool for statutory construction.

Brief for the Petitioner, *supra* note 24, at 3 (emphasis omitted).



precisely because they recognized that the explanation for the firing was *not* merely due to that individual's sex features, much less merely due to her being "overly aggressive."<sup>28</sup> What makes the act discriminatory is the relevance of acting on the basis of her being "overly aggressive" *in light of the social meanings of female sex*—particularly, the norm that females ought not to be aggressive. Therefore, we argue that instead of individual-level causal explanation, *social explanation* is better suited as a test for legal discrimination.

Nothing in the text or history of Title VII requires us to approach a discrimination case by asking about individual-level causation. Doing so simply confuses two different meanings of the word "because" in the statutory text's phrase: "because of . . . sex."<sup>29</sup> The sentence "V is injured *because* D drove into him" uses "because" in the individual-level causal sense.<sup>30</sup> The sentence, "Gideon wouldn't let Oona watch Game of Thrones because she is a child," by contrast, uses "because" in the sense of a social explanation. Here, the fact that Oona is a child explains Gideon's decision, but not in the sense that Oona's age *per se* explains his decision. Rather, the societal expectations and norms associated with Oona's age explain his decision. Such explanation is social: it identifies the cause of the event—here, Gideon's decision—at the level of social meanings and norms attached to the category *child*, rather than at the individual level of Oona's age.

Our positive argument is simply that Title VII's prohibition on sex discrimination "because of sex" calls for a social explanation, and not an individual-level explanation, for an instance of discrimination. A social explanation of an instance of sex discrimination asks whether the *social categories* that the individuals are members of, rather than individuals' sex features *per se*, explain the outcome in question. The full inquiry has three components. First, we need to determine the social meanings of sex categories. Second, we need to know if the complained-of act or practice is explained by something about those meanings. And finally, we need to know if it is wrongful in the way that Title VII prohibits. In a short essay, we can only gesture at how best to approach these three inquiries.

---

<sup>28</sup> See *Price Waterhouse v. Hopkins*, 490 U.S. 228, 250 (1989) ("In the specific context of sex stereotyping, an employer who acts on the basis of a belief that a woman cannot be aggressive, or that she must not be, has acted on the basis of gender.").

<sup>29</sup> Civil Rights Act of 1964, Pub. L. No. 88-352, tit. VII, § 703, 78 Stat. 255 (codified as amended at 42 U.S.C. § 2000e-2 (2018)). Conflating social explanation and individual-level causation is not limited to the law; we see the same practice in social science and en vogue causal modeling in computer science. For an example of a discussion on this topic, see Lily Hu & Issa Kohler-Hausmann, *What's Sex Got to Do with Fair Machine Learning?* (unpublished manuscript) (on file with authors). See also Amanda Shanor, *Sex Discrimination Behind the Veil Is Still Sex Discrimination*, TAKE CARE (Oct. 11, 2019), <https://takecareblog.com/blog/sex-discrimination-behind-the-veil-is-still-sex-discrimination> [<https://perma.cc/B8GB-YNUL>].

<sup>30</sup> See *supra* notes 15–17 and accompanying text.

First, how do we determine the social meanings of the relevant social categories? That is, how do we know which generalizations and norms of presentation, behaviors, roles, affects, etc. are associated with the categories “male” and “female”? This is a complex sociological question, but one that is essential to answer before we can know if the act/practice was because of sex. Interestingly, here is a place where counterfactual thought experiments are useful. When we imagine changing someone’s sex category, holding constant a *specific* presentation, behavior, role, or affect, and then conclude that the act/practice would have turned out differently, we learn that the features held constant are sexed—that is, they are evaluated in light of sex’s social meanings. But these tests cannot be mechanically applied because it is possible that toggling an employee’s sex would not change the outcome, and that this outcome was nevertheless “because of . . . sex.” (Imagine, for example, an employer who fires both male and female employees who display the qualities of being emotive, sensitive, and caring at the workplace, because he devalues qualities he deems “effeminate”). There will always be some armchair sociology and anthropology involved in this first inquiry. But it is important to be clear that we are doing armchair social science about the social meanings and relations of sex categories, and not about causal explanations.

Second, once we know the social meanings of sex categories, we need to know if the complained-of act/practice is explained by these meanings. To say that the outcome must be “explained by” these meanings could entail various relations. It could mean that a decisionmaker appeals to something about the category to justify his reason for action, or that generalizations about the group explain patterns of opportunities and selection. For example, in *Price Waterhouse* the employer appealed to the fact that the plaintiff cursed and did not wear makeup as a reason for denying her a promotion,<sup>31</sup> and in *R.G. & G.R. Harris Funeral Homes*, the employer explicitly pointed to presentation as the justification for the termination.<sup>32</sup> But in these cases, the employers did not object to “not wearing makeup” or “wearing skirts” *as such*, but rather objected to these things for some types of people. In other words, they took these features to be reasons for action *because* they were deemed inappropriate, given the norms associated with the relevant sex category. Alternatively, an unfavorable employment outcome might be “explained by sex” in the sense that general truths concerning the arrangements, roles, or practices associated with a sex category cause the complained-of policy to

---

<sup>31</sup> *Price Waterhouse*, 490 U.S. at 235.

<sup>32</sup> Brief for the Petitioner, *supra* note 24, at 1 (“Stephens handed [her employer] a letter . . . announc[ing] that Stephens had decided to start presenting and dressing as a woman at work . . . . In the end, [the employer] could not agree to Stephens’s plan to violate the dress code, so he offered Stephens a severance.”).

produce that outcome.<sup>33</sup> For example, a prohibition on promoting employees who have never missed work or been late will cause women to be disadvantaged, due to social-level arrangements and roles that assign more caregiving work to female persons.

Lastly, just as but-for causation can narrow potential parties to a tort action, but not assign liability, social explanation can tell us when the social meanings of sex categories explain an outcome, but cannot alone tell us if this outcome is *wrongful*, or, more specifically, *discriminatory*. What one needs for this final step is an independent normative theory about which social meanings are nefarious and contribute to systematic inequality. While we don't have space to defend such a theory here, we emphasize that one cannot escape the need for *some* substantive principle to distinguish between when an outcome is wrongful and when it is not. One cannot get away with saying that any outcome that is explained by the social meanings of a sex category is discriminatory without saying more about why this is so.<sup>34</sup> And this requires a normative principle—in particular, one that recognizes that any norm against discrimination on the basis of a given social category must be seeking to remake the social meanings associated with that category in *some* way. Why else would we prohibit discriminating on the basis of sex, but not shoe size, if we did not think that there was good reason to interrupt the reproduction of certain generalizations, stereotypes, and norms associated with the categories “male” and “female”? There is no way to determine which of these social meanings should be remade in the absence of substantive normative commitments. While one person might think that, for example, sex-specific dress codes reinforce nefarious meanings of sex, another might think that these social meanings are innocuous. Both positions are consistent with a social explanation of why a masculine-presenting woman was fired, but they differ as to whether this action counts as *discriminatory*.

---

<sup>33</sup> A careful reader might notice that an implication of our position is the collapse of disparate impact and treatment, as both are instances where the category “explains” the outcome. We cannot expand, much less defend, that position in this short paper. But we will point out that it is an upshot of our ontological position, which is that whenever one is acting on sex as such (say, by categorically excluding all people sex-coded female), one is acting on a category that is constituted by social meanings and relations, and when one is acting on one of the constitutive meanings or relations of sex, one is acting on sex. Therefore, there is no distinction between disparate impact and treatment to be had on the basis of the former acting on sex per se, and the latter acting on something that is not sex but merely correlated by it. See Issa Kohler-Hausmann, *Eddie Murphy and the Dangers of Counterfactual Causal Thinking About Detecting Racial Discrimination*, 113 NW. U. L. REV. 1163 (2019) (arguing that one cannot ground the ostensible distinction between disparate treatment and impact discrimination by appealing to acts/practices caused by race alone versus those caused by something that is not-race, but merely correlated with it, unless one holds a biological view about the category ‘race’).

<sup>34</sup> For example, the entire debate about affirmative action is best understood as a debate about whether or not it is fair and just to engage in some kind of admission rule that is unquestionably “explained by” sex or race.

It is here we get to the real heart of the matter in the pending *Zarda*, *Bostock*, and *R.G. & G.R. Harris Funeral* cases: which social meanings of sex categories ought Title VII disrupt? The defendants understand better than anyone that if norms and expectations about things like sexuality and gender presentation change, the meaning of sex in our society will change.<sup>35</sup> Precisely what the courts must decide, then, is which of the limits, expectations, norms, and roles imposed on the basis of sex classification ought to be tolerated and which ought to be changed. The statutory text does not give any guidance on this normative question. But we ought to be honest that this question—and not a metaphysical question about causality—is what we fundamentally are debating in these cases.

---

Preferred Citation: Robin Dembroff et al., *What Taylor Swift and Beyoncé Teach Us About Sex and Causes*, 169 U. PA. L. REV. ONLINE 1 (2020), <http://www.pennlawreview.com/online/169-UPa-L-Rev-Online-1.pdf>.

---

<sup>35</sup> For example, the defendants in *Altitude Express v. Zarda* argue that acceptance of the employees' arguments would create

Title VII protection whenever employees refuse to “conform” to what they consider to be “a normative sex-based stereotype.” [This] would empower employees who reject sex-based norms—even well-established, non-invidious ones—to antagonize employers. Male attorneys may insist on “wear[ing] nail polish and dresses” to court hearings; female swim instructors may put on fake beards or “strip to the waist” at work; and their employers would be helpless to stop them.

Brief for Petitioners, *supra* note 22, at 57. The defendants also noted that

adopting the analytical underpinnings of Zarda’s argument will revolutionize the meaning of sex discrimination in the workplace and produce staggering, indefensible outcomes[,] . . . overthrow[ing] . . . sex-specific policies for determining access to living facilities, sleeping quarters, restrooms, showers, and locker rooms; fitness tests for police, fire, and similar positions; and organizational dress and grooming standards.

*Id.* at 54-55. See also Brief for National Organization for Marriage and Center for Constitutional Jurisprudence as Amicus Curiae at 11-12, *Bostock v. Clayton County*, No. 17-1618 (U.S. Aug. 22, 2019) (“[T]he gender identity claimants are therefore seeking not just a minor adjustment to the civil rights laws, but a fundamental shift in policy and rejection of ‘common sense [and] decency’ that is inherent in the judicially-recognized fundamental right to bodily privacy from observation by persons of the opposite sex.” (quoting *Sepulveda v. Ramirez*, 967 F.2d 1413, 1416 (9th Cir. 1992))).