SUPREME CONFUSION ABOUT CAUSALITY AT THE SUPREME COURT

ISSA KOHLER-HAUSMANN (Yale Law School**) &**
ROBIN DEMBROFF (Yale Philosophy**)***

* We thank Erwin Chemerinsky, Jessica Clark, Ben Eidelson, Lily Hu, Andy Koppelman, Johnathan Schaffer, Scott Shapiro, John Witt, Gideon Yaffee for helpful engagement and comments.

Twice in the 2020 term, in *Bostock* and *Comcast,* the Supreme Court doubled down on a particular interpretation of how "but-for causation" applies to antidiscrimination statutes. According to the Court's reasoning, an outcome is discriminatory because of some status—say, sex or race—if the outcome would not have occurred "but-for" the plaintiff's status. We think this reasoning embeds profound conceptual errors that render the decisions deeply confused. Furthermore, those conceptual errors tend to limit the reach of antidiscrimination law. In this essay, we first unpack the ambiguity of the Court's interpretation and application of but-for causal reasoning. We then show that this reasoning arises from a misapplication of tort law principles within antidiscrimination law, and that, as a result, it is both conceptually and normatively indeterminate.

INTRODUCTION

In two 2020 cases, the United States Supreme Court interpreted two antidiscrimination cases to impose a particular "but-for causation" standard. The first case, *Comcast Corp. v. National Association of African American-Owned Media,* concerned what a plaintiff must plead under the § 1981 of the Civil Rights Act of 1866.[1] Section 1981 provides in relevant part that "persons . . . shall have the same right, in every State and Territory in the United States . . . to make and enforce contracts . . . as is enjoyed by white citizens."[2] The second case, *Bostock v. Clayton County, Georgia* (consolidated with *Altitude Express v. Zarda* and *R.G. & G.R. Harris Funeral Homes Inc. v. Equal Employment Opportunity Commission*), concerned whether Title VII's prohibition of discrimination

[1] Comcast Corp. v. Nat'l Ass'n Afr. Am.-Owned Media, 140 S. Ct. 1009, 1013-1015 (2020).
[2] Civil Rights Act of 1866, 42 U.S.C.A. § 1981(a) (West 2021).

"because of . . . sex" makes it unlawful to fire someone for being transgender or gay.[3] In both cases, the Court held that plaintiffs pursuing a claim under these statutes must plead and eventually show that the complained-of outcome would not have occurred "but-for" the plaintiff's sex or race—that is, if the plaintiffs' sex or race were different than it in fact is.[4]

The *Comcast* and *Bostock* decisions suggest that this causal test—which we'll call the "but-for causal test"—defines discrimination. More specifically, they suggest that if an outcome would not have occurred but-for a plaintiff's protected status (e.g., as a woman, as Black), then the outcome is an instance of discrimination.[5]

In this paper, our broad hope is to persuade you that this understanding of discrimination embeds profound conceptual errors that render the decisions deeply confused, and that also threaten to limit the reach of antidiscrimination law. What makes a given outcome discrimination, we think, is not whether it was *caused* by a plaintiff's status. Instead, as we will show, it is a question of whether it was caused by an action or policy that acted upon or reinforced nefarious social meanings associated with a protected status or hierarchical social positions that are typical of these statuses. However, our goal in this paper is not to defend a full theory of discrimination. It is, rather, to show that one is still needed—the but-for causal test alone cannot tell us what discrimination is.

We make two interrelated arguments in this paper. First, the causal showing that the Supreme Court articulates in *Bostock* and *Comcast*—that the plaintiff show that the outcome would not have occurred but-for the plaintiff's sex or racial status—cannot serve as an independent test of what counts as discrimination because it is inherently indeterminate. One cannot set up the counterfactual thought experiment without more specificity about the relevant counterfactual contrasts, and one cannot choose which counterfactual contrasts are relevant without a prior normative theory of how the defendant ought to have behaved vis-a-viz the plaintiff's sex or racial status. Second, the relation of counterfactual dependence alone cannot give us such a normative theory. Rather, any causal test in antidiscrimination law must rest on a prior and independent

---

[3] *See* Bostock v. Clayton Cnty., Ga., 140 S. Ct. 1731, 1737-39 (2020).

[4] *Compare Comcast Corp.*, 140 S. Ct. at 1019 ("To prevail, a plaintiff must initially plead and ultimately prove that, *but-for race*, it would not have suffered the loss of a legally protected right."), *with Bostock*, 140 S. Ct. at 1739 ("So long as the plaintiff's sex was one but-for cause of that decision, that is enough to trigger the law.").

[5] *See Bostock*, 140 S. Ct. at 1748 ("When a qualified woman applies for a mechanic position and is denied, the 'simple [but for] test' immediately spots the discrimination: A qualified man would have been given the job, so sex was a but-for cause of the employer's refusal to hire."); *see also infra* Parts II-B and II-C.

theory of what people are owed in various domains (e.g. employment, contracts, etc.) *given* the social meanings and relations of sex or race in our society.

Part I of this essay focuses on unpacking the inherent ambiguity of the but-for causal test and the problems this ambiguity can cause within antidiscrimination law. This test, we point out, relies on counterfactual questions—that is, asking questions about alternative ways things might have been. But what alternative ways do we imagine? Which of the many alternative possibilities we choose to entertain determines the outcome of the but-for causal test, but the Court leaves these alternatives ambiguous thereby creating a deep indeterminacy in its decision-making.

Within tort law, as we point out in Part II, the range of relevant alternative possibilities are constrained by the question of duty. Namely, liability attaches only in those case where damages would not have occurred if a *negligent* or *duty-defying* action or policy had not occurred. This, we note, is strikingly different from the question of whether damages would have occurred if the action or policy *per se* had not occurred. This difference is essential to our primary criticism of the Court's applications of the but-for test within antidiscrimination law. Were the Court's applications to mirror those in tort law, the relevant question for the Court would be whether certain damages (e.g., losing a job or promotion) would have occurred if a *discriminatory* action or policy had not occurred. But this is not what we find. Instead, we find the Court asking whether damages would have occurred if the plaintiff had a different *status* (e.g., a different race or sex). This is then coupled with the idea that, if the answer is "no," this shows that an action or policy that caused the alleged damage was discriminatory. This is not only a misapplication of torts doctrines, but also a misapplication that renders these decisions hopelessly confused, and that potentially limits the reach of antidiscrimination law.

In Part III, we focus on an alternative way to use the but-for test within antidiscrimination law. With this approach, a court must first determine whether an action or policy is discriminatory and only then—assuming the answer is "yes"—ask whether damages would have occurred but-for that discriminatory action or policy. This would make applications of the but-for test within antidiscrimination law mirror its applications within tort law, where a substantive theory of duty frames which causal questions are relevant to the legal inquiry. It would also address the counterfactual test's inherent ambiguity. However, this alternative would require that the Court makes explicit what is currently hiding beneath the Court's talk of causation: a substantive theory of what counts as wrongful or unequal treatment in light of or in virtue of (i.e., *because of*) the

social categories of race, sex, religion, etc. A substantive theory of how people ought to be treated given how social categories are structured is a predicate question before deciding whether a plaintiff suffered from discrimination. The relation of causal dependence alone cannot deliver such a theory. While giving a full account of discrimination is beyond the scope of this paper, we suggest that a good way forward is for courts to consider whether damages were caused by an action or policy that acted upon or reinforced nefarious social meanings associated with a protected status (e.g., stereotypes of women) or hierarchical social positions that are typical of these statuses (e.g., the subordination of women).

This essay will not focus on questions of statutory interpretation, such as how to fix the semantic content of a statute. We do not intend to settle which is the right interpretation of the written law, but we do think that any plausible candidate must offer a coherent way to identify discrimination and sensical application of causation principles. If some methodology compels an interpretation that relies on confused use of a but-for test to define discrimination, we think that counts against the methodology. In addition, and to streamline our analysis, we will focus on *Bostock* and sex discrimination, and largely limit our analysis of *Bostock* to the case of firing an employee for being gay. However, we believe our analysis equally applies to the case of firing an employee for being transgender, as well as to other cases of discrimination under Title VII, such as race discrimination and religious discrimination.

I. THE INHERENT AMBIGUITY OF THE BUT-FOR CAUSAL TEST

The Supreme Court has long interpreted antidiscrimination statutes through the lens of "but-for" causation, borrowed from tort law.[6] One of the foremost understandings of causation in philosophy is cashed out in terms of a

[6] *See, e.g.*, Charles A. Sullivan, *Tortifying Employment Discrimination*, 92 B.U. L. Rev. 1431, 1436 (2012); Michael C. Harper, The Causation Standard in Federal Employment Law: Gross v. Fbl Financial Services, Inc., and the Unfulfilled Promise of the Civil Rights Act of 1991, 58 Buff. L. Rev. 69, 83 (2010); Mark S. Brodin, *The Standard of Causation in the Mixed-Motive Title VII Action: A Social Policy Perspective*, 82 Colum. L. Rev. 292, 312 (1982). The Court is explicit about drawing on tort doctrines. *See, e.g.*, Staub v. Proctor Hosp., 562 U.S. 411, 417 (2011) ("[W]e start from the premise that when Congress creates a federal tort it adopts the background of general tort law"). But as we argue this alone does not determine which of the many possible formulations of the causal showing one endorses.

counterfactual, expressed as "a conditional sentence in the subjunctive mood."[8] To create a counterfactual conditional of this sort, you must first propose an antecedent that is contrary to actual fact, and then consider what would have come about under these contrary-to-fact conditions.[9]

For example, suppose you have just dropped a water glass, and the glass shattered on the kitchen floor. One potential counterfactual question you might pose is:

> If I had not dropped the glass, would it have shattered on the kitchen floor?

Now, suppose you answer "no." This means that you think that dropping the glass was a "but-for" cause of the glass shattering: the glass wouldn't have shattered *but-for* the fact that you dropped it. Sounds simple, we know.

But there is an important and inherent ambiguity in the analysis of these counterfactual questions. Notice that we haven't actually specified what else you might have done with the glass if not drop it. To see why specifying the counterfactual contrast matters for producing a determinate answer to causal questions, it is helpful to consider the philosopher David Lewis's "possible worlds" approach to analyzing counterfactual conditionals.[10] A possible world is a complete, alternative reality, or what Christopher Menzel calls a "single, maximally inclusive, all-encompassing situation."[11] There are as many possible worlds as there are ways that things might have been.[12] If you think that you could have one less hair on your head, or have chosen not to read this article, or have never learned to read at all, then there are possible worlds where those things are true.[13]

---

[8] John Collins et al., *Counterfactuals and Causation: History, Problems, and Prospects*, *in* CAUSATION AND COUNTERFACTUALS 1, 2 (JOHN COLLINS ET AL. EDS., 2004).
[9] *Id.* at 2-3.
[10] *See generally* David Lewis, *Counterfactuals and Comparative Possibility*, 2 J. PHILOS. LOG. 418 *passim* (1973) [hereinafter Lewis, *Comparative Possibility*]; David Lewis, *Counterfactual Dependence and Time's Arrow*, 13 NOÛS 455 (1979) [hereinafter Lewis, *Time's Arrow*]; *see also* Collins et al., *supra* note 8, at 455.
[11] Christopher Menzel, *Possible Worlds*, STAN. ENCYC. OF PHIL., https://perma.cc/7BM4-QFVR (last updated Sept. 21, 2021).
[12] *Id.* ("[T]hings might have been different in countless ways, both trivial and profound. History, from the very beginning, could have unfolded quite other than it did in fact."); Lewis, *Comparative Possibility*, *supra* note 10, at 420 ("Differences never come singly, but in infinite multitudes.").
[13] Strictly speaking, Lewis would say that the person in the possible worlds is not you, but your "counterparts"—i.e., people who resemble you, but are not identical to you. *See* David Lewis, *Counterparts of Persons and Their Bodies*, 68 J. PHIL. 203, 203-11 (1971).

We can appeal to what would happen in relevant possible worlds to analyze if a particular counterfactual is true. Return to the question of whether dropping the glass caused the glass shattering. Using Lewis's framework, we would start by considering all possible worlds in which you did *not* drop the glass. We would then ask, "Is it true in all, some, or none of these worlds that the glass shatters?" We hope you share our intuition that the answer is "some." After all, there are possible worlds where you did not drop the glass, but instead shattered it by squeezing too tightly, or where you instead threw the glass against the wall. In those worlds, the glass still would have shattered, showing that the dropping is not a but-for cause of its shattering in *all* possible worlds.

Put simply, the counterfactual question, "If I had not dropped the glass, would it have shattered on the kitchen floor?" leaves ambiguous which worlds are relevant to consider when answering this question. This ambiguity is typical of counterfactual questions—as Lewis wrote, "[c]ounterfactuals are infected with vagueness, as everyone agrees."[14] To answer our counterfactual question by appeal to what would happen in possible worlds, we must first delimit *which* possible worlds are relevant for our analysis by further disambiguating our question.[15] Otherwise, the question is hopelessly indeterminate. In everyday conversations, this delimiting often implicit or assumed. We often know simply from context what possibilities we are being asked to consider. In the example above, most likely we just want to know what would have happened if you had not dropped but *held* the glass and did so in a kitchen like your actual kitchen, where the laws of nature apply are as they actually are, and so on. In other words, there is often tacit agreement about which alternative possibilities are relevant in the context of everyday questions about causation.

There is no such tacit agreement, however, in the context of antidiscrimination law. Instead, ambiguity regarding which worlds are relevant to the counterfactual question runs rampant. Courts and commentators disagree about which counterfactual contrasts—i.e. which possible worlds—should be considered when analyzing what would have occurred "but-for" a plaintiff's social status such as sex.[16] Therefore, they come to different answers about

---

[14] Lewis, *Time's Arrow*, *supra* note 10, at 457.
[15] Lewis, *Comparative Possibility*, *supra* note 10, at 418-20.
[16] For examples, the Defendants and Trump Department of Justice argued that the appropriate counterfactual contrast for the a gay male plaintiff would be a female lesbian, Brief for Petitioners at 9, Altitude Express, Inc. v. Zarda, No. 17-1623 (U.S. Aug. 16, 2019); Brief for the United States as Amicus Curiae at 19, Bostock, No. 17-1618 (U.S. Aug. 23, 2019), whereas the Plaintiffs proposed that the appropriate counterfactual contrast would be a straight female, Transcript of Oral Argument at 7-8, Bostock v. Clayton County, No. 17-1618 (U.S. argued Oct. 8, 2019).

whether sex was the cause of the firing because they are actually asking different counterfactual questions. To see what we mean, let's take a closer look at *Bostock*. *Bostock* presented the question of whether discharging workers because they are transgender or gay counts as discrimination "because of sex" under Title VII.[17]

The Majority opinion, authored by Neil Gorsuch, answered "yes."[18] The logic of this opinion begins with the basic legal rule: under Title VII, firing someone because of sex is unlawful discrimination.[19] From here, three assumptions are added:

(A) The phrase "because of…sex" should be interpreted to mean but-for sex.[20]

(B) Firing someone but-for a status that is defined in part by their sex entails firing them but-for their sex.[21]

(C) An individual's status as gay or transgender is defined in part by sex.[22]

Putting these pieces together, Gorsuch reasons that, because the employees would not have been fired in an alternative world where they are not gay or transgender, these employees were targets of unlawful discrimination.[23]

We can formalize this as follows, where the arrows (or absence) indicate but-for causal relations (or lack thereof), and "P" stands for plaintiff.

---

[17] *Bostock*, 140 S. Ct. at 1737 (2020) ("Today, we must decide whether an employer can fire someone simply for being homosexual or transgender [under Title VII].").

[18] *Id.* at 1741-43 (2020) ("For an employer to discriminate against employees for being homosexual or transgender, the employer must intentionally discriminate against individual men and women in part because of sex.").

[19] *Id.* at 1740 ("[A]n employer who intentionally treats a person worse because of sex—such as by firing the person for actions or attributes it would tolerate in an individual of another sex—discriminates against that person in violation of Title VII.").

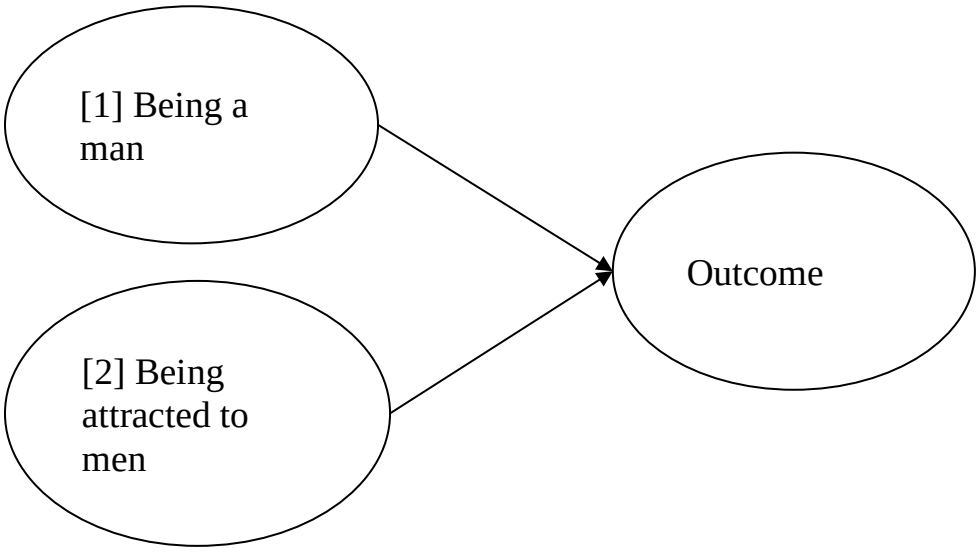
[20] *Id.* at 1739 ("Title VII's 'because of' test incorporates the '"simple"' and 'traditional' standard of but-for causation.") (quoting Univ. of Tex. Sw. Med. Ctr. v. Nassar, 570 U.S. 338, 346, 360 (2013))

[21] *Id.* at 1741 ("If the employer intentionally relies in part on an individual employee's sex when deciding to discharge the employee—put differently, if changing the employee's sex would have yielded a different choice by the employer—a statutory violation has occurred.").

[22] *Id.* ("[I]t is impossible to discriminate against a person for being homosexual or transgender without discriminating against that individual based on sex.").

[23] *See id.* at 1741-43 ("For an employer to discriminate against employees for being homosexual or transgender, the employer must intentionally discriminate against individual men and women in part because of sex..").

Figure 1: The Majority's Logic



The Majority's (and Dissent's) first interpretive move is to assert that Title VII requires the plaintiff to show that the complained-of outcome would not have happened but-for the plaintiff's sex (A). [24] But, "as everyone agrees,"[25] that question is infected with vagueness. Which counterfactual worlds are relevant to analyzing whether the but-for causal showing demanded is true? The Majority's second assumption conceptually separates "being gay" into [1] the employee's sex and [2] the sex of the employee's real or imagined sexual partners.[26] With this separation, the Majority reasons that, to answer the counterfactual question, we should consider alternative worlds where we change [1] and hold [2] fixed—that is, we should ask whether a gay employee would have been fired in a possible world where he was instead a straight woman, or whether a lesbian would have been fired in a possible world where she was instead a straight man.[27] On the

[24] *Id.* at 1742: "If an employer would not have discharged an employee but for that individual's sex, the statute's causation standard is met, and liability may attach."
[25] Lewis, *Time's Arrow*, *supra* note 10, at 457.
[26] *Id.* at 1742 ("[T]wo causal factors may be in play—*both* the individual's sex *and* something else (the sex to which the individual is attracted or with which the individual identifies).").
[27] *See, e.g.*, *id.* ("A model employee arrives [at an office party] and introduces a manager to Susan, the employee's wife. Will that employee be fired [under a company policy mandating discharge of employees known to be homosexuals]? . . . [T]he answer depends entirely on whether the model employee is a man or a woman.").

plausible assumption that an employer would not have fired the gay plaintiff if he were a straight woman, the Majority concludes that the firing was discriminatory.[28]

However, this logic is wide-open to the types of criticisms presented in Justice Alito's dissent, showing that choosing a different counterfactual contrast can yield a result suggesting that sex discrimination did *not* occur. Justice Alito's dissent can be read as a disagreement with the Majority's way of delimiting the possible worlds relevant to the establishing the but-for causation they both hold to be essential for liability under Title VII. The Majority decision, Alito correctly observes, rests on conceptualizing being gay as being a man attracted to man or a woman attracted to woman, as opposed to being a person who is same-sex attracted.[29] It is only because of this conceptualization that the Majority can proceed to operationalize the vague causal question of (a1)—Did the firing happen because of the plaintiff's sex?—in terms of the specific counterfactual in Figure 1—toggle [1] and hold [2] fixed.

Alito thinks that, if an employee is fired because he is gay, it is a slight of hand to conceptualize being gay as a man attracted to men—a framing that lends itself to the showing that he was fired but-for his sex.[30] Instead, Alito argues, the employee was fired because he is same-sex attracted, meaning that the relevant counterfactual is not whether he would have been if he were a straight woman, but instead whether he would have been fired if he were a gay woman.[31] (He assumes the answer is "yes.")[32] In other words, rather than hold the plaintiff's sexual attraction fixed while toggling their sex, Alito argues that we should hold their same-sex attraction fixed and *then* toggle their sex.[33] This, he claims, shows that the plaintiffs' unprotected status as gay, and not their protected status as men or women, caused them to be fired.[34]

---

[28] *Id.*

[29] *Id.* at 1762 (Alito J., dissenting) ("The Court [Majority] tries to avoid this inescapable conclusion by arguing that sex is really the only difference between the two employees. This is so, [they] maintain[], because both employees 'are attracted to men.' Of course, the employer would couch its objection to the man differently. . . . his sexual orientation.") (internal citations omitted).

[30] *See id.* (Alito J., dissenting) ("[T]he Court loads the dice. . . . because in the mind of an employer who does not want to employ individuals who are attracted to members of the same sex, these two employees are not materially identical in every respect but sex [because one is gay and one is straight].").

[31] *Id.* at 1763 (Alito J., dissenting) ("It can easily be shown that the employer's real objection is not 'attract[ion] to men' but homosexual orientation.").

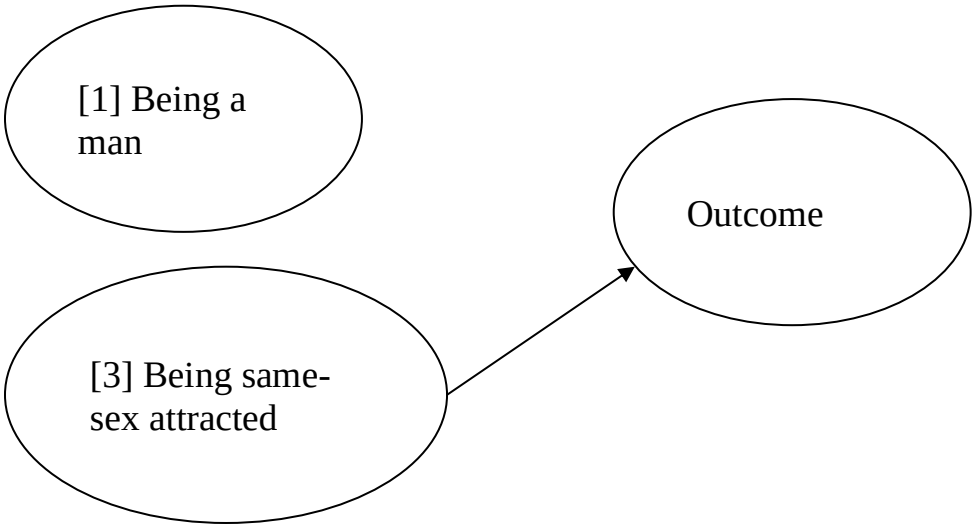
[32] *See id.* (Alito J., dissenting) ("It is attraction to members of their own sex—in a word, sexual orientation. . . . we can infer, is the employer's real motive.").

[33] *Id.* (Alito J., dissenting).

[34] *Id.* at 1762 (Alito J., dissenting) ("Title VII allows employers to decide whether two employees are 'materially identical.' Even idiosyncratic criteria are permitted; if an employer thinks that

Let's again formalize this logic:

Figure 2: Dissent



The stalemate between Figure 1's and Figure 2's operationalization of the but-for sex causal question—what Alito calls a "battle of labels"[35]—indicates an important shortcoming of using but-for causation as a litmus test for discrimination. Using the Court's but-for test, we must answer whether or not sex "caused" the outcome. But to do so, *you have to delimit the relevant possible worlds for the counterfactual*, meaning you must designate what to change and what to hold fixed.[36] And which counterfactual worlds we entertain determines whether or not we find that the relevant outcome (e.g., the firing) would have obtained. The Majority holds that the employee's sex was a but-for cause of being fired because they only entertain worlds where the biological sex and sexual orientation of the plaintiff have been changed (i.e. lesbian compared to a straight

Scorpios make bad employees, the employer can refuse to hire Scorpios. Such a policy would be unfair and foolish, but under Title VII, it is permitted.").

[35] Alito characterizes this as a "battle of labels," meaning "[i]f the employer's objection to the male employee is characterized as attraction to men, it seems that he is just like the woman in all respects except sex . . . . On the other hand, if the employer's objection is sexual orientation or homosexuality, the two employees differ in two respects." *Id.* In an astounding act of circular hypocrisy, Alito at once acknowledges the inherent indeterminacy of the counterfactual test, but nonetheless asserts that he produces the right answer using the counterfactual test because his contrast is the right one: "[H]owever, there is no standoff. It can easily be shown that the employer's real objection is not 'attract[ion] to men' but homosexual orientation." *See id.* at 1763.

[36] *See generally supra* notes 10-15 and accompanying text.

man, or gay man compared to a straight woman). By contrast, Alito holds that the employee's sex is *not* a but-for cause because he only entertains worlds where the biological sex, but not the sexual orientation, of the plaintiff have been changed (i.e. lesbian compared to gay man).

The takeaway, is that once we take "because of sex" to mean "sex as a but-for cause", there is nothing in either the text of Title VII or the nature of causation that requires any particular view as to which counterfactual possibilities are the relevant ones to consider.[37] Should we consider possible worlds where we toggle the plaintiff's biological sex but hold constant the plaintiff's male-sex sexual attractions (as the Majority proposes), or possible worlds in which we toggle the plaintiff's biological sex and hold constant the plaintiff's same-sex sexual attraction (as the Dissent proposes)? Embracing Lewis's possible worlds approach to analyzing counterfactuals does not resolve this conundrum. The possible worlds approach directs analysis to what would have happened in worlds that are *relevant* to the question being asked.[38] But the relevant question just *is* the question of what counts as discrimination because of sex, the precise question to which—the Court imagines—the counterfactual test will deliver an answer!

Said yet another way, to assert that this or that possible world is "relevant" to a particular causal inquiry is just another way of stating what one takes the guiding question to be. In the previous example of the dropped glass counterfactual question, ordinary speakers typically agree about which possible worlds are relevant to answering the question because they agree on the precise question implicit in a particular semantic articulation (i.e., "If I had not dropped the glass, would it have shattered on the kitchen floor?" means "What would happen if I held instead of dropping the glass, in a kitchen like my actual kitchen, where the laws of nature apply are as they actually are, and so on…"). In the context of antidiscrimination law, however, people disagree about the normative and legal definition of unlawful discrimination because of sex. That is, of course, the premise of this entire inquiry. Therefore, they disagree about the precise counterfactual question being asked.

As a result, they disagree about what should be held fixed in the background of the analysis—i.e., which possible worlds are *relevant* for analyzing this question. Like the Court, we can adopt the vague formal definition of

---

[37] Robin Dembroff et al., *What Taylor Swift and Beyoncé Teach Us About Sex and Causes*, 169 U. PA. L. REV. ONLINE 1, 7 (2020), https://perma.cc/WL8E-CWGE.

[38] Lewis, *Comparative Possibility*, *supra* note 10, at 419–20 ("In considering the supposition 'if I had just let go of my pen... [sic]' I will go wrong if I consider bizarre worlds [with an anomalous] law of gravity . . . whereas in considering the supposition 'if the planets traveled in spirals... [sic]' I will go just as wrong if I ignore such worlds.").

discrimination as be being treated worse than others similarly situated in all relevant respects, but there is nothing in Title VII that tells us *what are* these relevant respects, or at what level of abstraction to describe them.

That is, do we describe the male plaintiff in *Bostock* as being similarly situated to (some) women in the workplace at the level of abstraction the Majority proposes, i.e. as having male-sex sexual attractions? Or do we describe him as being similarly situated to (some) women in the workplace at the level of abstraction as the Dissent proposes, i.e. as having same-sex sexual attraction?[39] Nothing about the metaphysics of causality tells us which question to pose. And yet, the legal question of labially turns on the answer because the defendant is treated worse than those similarly situated under the first formulation, and not under the second. Therefore, even if we adopt the Court's but-for reading of Title VII, then, we left at an impasse regarding whether a plaintiff's sex was the "but-for" cause of a particular outcome.

II. CASUAL CONFUSION: TAKING FROM TORTS, MISUSING FOR DISCRIMINATION

So far, we have focused on the inherent ambiguity that attends the question of whether a given outcome would have occurred but-for a plaintiff's protected status, such as sex or race. In this second part, we want to pose an even deeper problem for the use of the but-for causal test within antidiscrimination law. This is a problem with the more basic assumption that the question, "Did discrimination occur because of sex?" can and should be translated into the question, "Did the outcome occur but-for the plaintiff's sex?" This translation, we show, is a distortion of the but-for causal doctrine found in tort law and sets up the absurd idea that we can determine what is discrimination—a deeply normative notion—simply by looking at causal relations.

A. The But-For Test in Tort Law

[39] Notice, this is the identical relational feature about the actual plaintiff, just described at two different levels of abstraction. Neither is a more "accurate" description of the plaintiff's relational feature, nor of what was salient to the defendant's decisionmaking. The salient feature about the plaintiff for the defendant's decision was the *relation* of the plaintiff's sex to the sex of his real or presumed sexual attraction. For *this* plaintiff (who was male), that identical feature can be described at a lower level of abstraction (male-sex sexual attractions) or at a higher level of abstraction (same-sex sexual attractions).

But-for causal tests are central to tort law.[40] Within tort law, though, the ambiguity inherent to these tests is ameliorated by the structure of the legal question at hand.[41] The question, in cases of tort law, is not simply whether certain damages would have occurred but-for the defendant's actions; instead, courts ask whether those damages would have occurred but-for the defendant's *wrongful, duty-defying,* or *negligent* actions.[42] In other words, the use of the but-for causal test within tort law is set up not to tell us whether the defendant's behavior is the cause of the plaintiff's loss, but instead to tell us whether their *duty-defying* behavior is the cause of the plaintiff's loss.[43]

By setting the question up in the way, courts are given a structure that helps them delimit which counterfactual scenarios are the relevant ones for determining the defendant's liability. To see what we mean, consider the classic tort case where the defendant ["D"] drives 65 mph in a 25-mph zone, does not break in time for a child ["V"] crossing the street in the crosswalk, and subsequently hits and injures V. From here, we might ask:

> Would V have suffered this injury if it weren't the case that D drove 65 mph on the road?

Our answer, of course, depends on what D *would have been doing* if not driving 65 mph. If we simply asked the above question, we would not know what we are supposed to imagine D doing *instead* of driving at 65 mph. Is D driving at 70 mph, not driving at all, driving at 25 mph, or igniting his car with a torch? These details matter. If we imagine that D was instead driving 70 mph, we might conclude that yes, V *would* have still suffered the injuries, suggesting that D's driving was *not* a but-for cause of V's injuries.[44] But if we imagine that D was

---

[40] *See Comcast*, 140 S. Ct. at 1014 (2020) ("It is 'textbook tort law' that a plaintiff seeking redress for a defendant's legal wrong typically must prove but-for causation.") (quoting Univ. of Tex. Sw. Med. Ctr. v. Nassar, 570 U.S. 338, 347 (2013)).

[41] *See, e.g.*, Jane Stapleton, *Choosing What We Mean by "Causation" in the Law*, 73 MO. L. REV. 433, 438-441 (2008); *see also* H.L.A. HART & TONY HONORÉ, CAUSATION IN THE LAW 133-36 (2d ed. 1985).

[42] *See* RESTATEMENT (THIRD) OF TORTS: LIAB. FOR PHYS. AND EMOT. HARM §§ 3, 6 (Am. Law. Inst. ed. 2010).

[43] *See id.* § 26.

[44] Even the distance from V at which we imagine changing D's speed is determined by these normative intuitions. For example, if at 25 mph one can safely break within the distance of 30 feet, then one needs 200 feet to safely stop at a rate of 65 mph. Now, assume that, in the actual event, V ran in front of D's car at a distance of 40 feet. The relevant counterfactual asks us to imagine that at the exact moment D was 40 feet from V, D was going 25 mph, and then ask if D would have been able to successfully break. Similarly, a counterfactual where D was going 70 mph for 10 minutes prior to the accident, and so drove past V by the time V ran into the street, would not be

instead driving 25 mph, we would likely conclude that no, V *would not* have suffered the injuries, suggesting that D's driving *was* a but-for cause of V's injuries.

As with the earlier shattered glass example, we again see that how we answer a counterfactual question depends on which possible worlds we consider relevant to our question. Within tort law, though, the relevant counterfactual would be widely understood as:

> If D *had not been driving negligently*, would V have suffered these injuries?[45]

In other words, we are asked to imagine that the defendant acted in a particular lawful or duty-conforming manner, and then—if the answer is 'no'—ask if the plaintiff's loss would still have occurred if D had acted as duty demanded.[46]

Here, we hope it is clear that the relevant but-for test requires a prior and separate specification of duty or some other functionally similar legal-normative concept to get off the ground. Without that specification, it is entirely unclear what D is imagined to be doing in our counterfactual thought experiment, or why the proposed causal dependence grounds D's liability for P's loss. These normative concepts circumscribe the class of relevant counterfactual contrasts (things D would be doing, if not driving 65 mph) that we can consider in establishing that D's action was a "but-for" cause of the outcome, and that D is thereby liable for V's loss.

In short, nothing in the abstract requirement that the plaintiff's loss be causally dependent upon the defendant's act (or culpable omission) tells us *which* (if either) of the above counterfactual set ups to consider, the one in which D is driving 70 mph or the one in which D is driving 25 mph. And yet, we would almost certainly choose the latter set-up, because we think that the *relevant* question is whether the defendant's breach of their legal duty to adhere to the speed limit is a but-for cause. Jonathan Schaffer explains these intuitions about the relevant counterfactual contrast as intuitions that are driven by our knowledge of what is lawful, duty-conforming conduct.[47] We agree. But this means that

---

the relevant counterfactual to entertain. Both the imagined speed and distance are constrained by our concept of what D *ought* to have been doing.

[45] *See* RESTATEMENT (THIRD) OF TORTS: LIAB. FOR PHYS. AND EMOT. HARM §§ 3, 6 (Am. Law. Inst. 2010).

[46] *See id.*

[47] Jonathan Schaffer, *Contrastive Causation in the Law*, 16 LEGAL THEORY 259, 272 n.31 (2010) ("The appeal to lawful conduct is hardly new. . . . [For instance, Stapleton] speaks of the law as

causal determinations within tort law necessarily depend on prior concepts such as duty and lawfulness, or some other legal-normative concepts that help us delimit the relevant possible worlds for us to consider.

None of what we have said suggests that the "but-for" causal test is useless or reduces to a normative judgement. For one thing, this test reveals instances where, even if the defendant *had* conformed to their duty, the plaintiff still would have suffered the relevant loss.[48] But still, the test cannot do this work without some prior, independent account of the content of D's duty to delimit which counterfactual contrasts are relevant to the legal question at hand.[49]

Imagine that a drug company has a duty to manufacture its bottles in a "childproof" fashion, but has failed to do so. Now let us assume that a child throws one of the company's non-childproof bottles at another child's face, causing injuries. We want to know whether the company is liable for these injuries. Here, we might be tempted to ask:

> If the drug company had not manufactured bottles, would the child have suffered these injuries?

The answer, someone might argue is "no"—if the bottle had not existed, the child would not have suffered their injuries. But, even if this is right, this doesn't tell us that the company is *liable* for those injuries. Without any normative framing, the but-for test only can tell us whether the company's actions are a but-for cause of the child's injuries, and not anything about legal liability. After all, the company is not under a duty to refrain from manufacturing bottles altogether. What we *really* want to know is this:

---

providing 'filtering devices' that 'specify relevant hypothetical comparator worlds' where 'the specified factor in turn determines the hypothetical worlds (because these are no-breach worlds).'").

[48] Consider the case summarized by Mark Brodin in which the "[p]laintiff brought an action for the wrongful death of his eighteen-month-old child. The complaint alleged that the infant was poisoned by ingesting pills distributed by the defendant drug company in a container that was not labeled 'poison' or marked with a skull and crossbones, as required by law." Later, the unmarked pills were ingested by a baby who "was too young to have understood any warning on the bottle even if it had been there." Mark S. Brodin, *The Standard of Causation in the Mixed-Motive Title VII Action: A Social Policy Perspective*, 82 COLUM. L. REV. 292, 314 (1982).

[49] We are not taking the "extreme" realist position that legal causal determinations are nothing more than normative determinations. *See generally* Joshua Knobe & Scott Shapiro, *Proximate Cause Explained: An Essay in Experimental Jurisprudence*, 88 U. CHI. L. REV. 165, 169 (2021). Rather, we are simply saying that *which* causal dependencies we ask about is a function of normative positions about what we owe each other. The relation of counterfactual factual dependence is still a matter of metaphysics, not ethics.

> If the drug manufacturer had made its bottles in conformity with its duty (i.e., as "childproof"), would its bottle have caused the child's injuries?

Here, it seems, the answer is "yes". The child's injuries would still have occurred if the bottle had "childproof" safeguards, and so the drug company is not liable to the plaintiff. Thus, the drug manufacturer's breach of duty was not a but-for cause of the child's injuries. This example shows that the relevant application of the but-for test in tort law requires a prior normative-legal concept in order to answer the relevant counterfactual question that determines liability.

In tort law, but-for causation also authorizes the plaintiff to haul *this particular defendant* into court for *these particular losses*.[50] If you allege in a tort claim that Bill Gates breached a duty of care while rolling around in his Bentley at 90 mph in California and that you were harmed by a passing car in New York, then you haven't stated a valid claim against Bill Gates. Even under a broad delimiting of which possible worlds are relevant, it is not true that your injuries would not have occurred but-for Gates's behavior. But, supposing Gates *had* hit you, our point is that this alone does not supply the normative grounds for assigning legal liability: doctrines of duty or other legal-normative assumptions about entitlements or lawful conduct do that.[51] In order for Gates to be liable for damages, you would need to show that Gates's unlawful or negligent behavior, and not simply his behavior *per se*, was the but-for cause of your damages.[52] The "but-for" causal test is not a test for what *counts* as duty or negligence. When

---

[50] *See* RESTATEMENT (THIRD) OF TORTS: LIAB. FOR PHYS. AND EMOT. HARM § 26 cmt. f (AM. LAW. INST. 2010); *see also* Ernest J. Weinrib, *Corrective Justice in a Nutshell*, 52 U. TORONTO L. J. 349, 352 (2002). But even if a plaintiff alleges a causal connection between the plaintiff's loss and defendant's breach of duty, there often are difficult issues of proof, and there are also doctrines and cases that stretch or even eliminate the causal requirement. *See, e.g.*, Sindell v. Abbott Lab'ys, 26 Cal.3d 588, 614-23 (1980) (Richardson, J., dissenting) (characterizing the majority's description of the doctrine of market share liability as eliminating the causation connection between the plaintiff and defendants as "expressly abandon[ing]" the "traditional requirement" of actual causation between the defendants' act and plaintiffs' resulting injury).

[51] A substantial body of literature on corrective justice is dedicated to debating these issues. *See, e.g.*, Jules Coleman, *The Mixed Conception of Corrective Justice*. 77 IOWA L. REV. 427, 439-40 (1992); Weinrib, *supra* note 50, at 352-54.

[52] *See* RESTATEMENT (THIRD) OF TORTS: LIAB. FOR PHYS. AND EMOT. HARM § 6 (AM. LAW. INST. 2010) ("An actor whose *negligence* is a factual cause of physical harm is subject to liability for any such harm within the scope of liability . . . .") (emphasis added); *id.* § 26 cmt. F ("The second framing step [of the but-for causal inquiry] requires determination of the conduct of the actor alleged to be tortious, which also entails identifying the alternative conduct that would not have been tortious.").

determining liability, the question of what duties or entitlements we have to each other must be answered *independently of* and *prior to* running this causal test.[53]

Put simply, a tort plaintiff must show that the defendant's breach of duty or unlawful action was a but-for cause of their damages, not the defendant's action per se. Framing the counterfactual contrast in these normative-legal terms gets at the heart of what is at issue in these cases, and helps courts treat the inherent vagueness in counterfactual questions by narrowing which possible worlds are relevant for answering the question.[54]

B. The But-For Test in Antidiscrimination Law

We turn now to the use of "but-for" causal test in the context of antidiscrimination law. Here, we see that the Supreme Court has long interpreted antidiscrimination statutes through the lens of "but-for" causation, borrowed from tort law.[55] In *Comcast*, for example, the Court cites precedent relying on a tort law treatise, reasoning that, "[t]his ancient and simple 'but for' common law causation

---

[53] *See id.* § 6 cmt. B ("The first element [of a prima facie case], duty, is a question of law for the court to determine . . . [and] [e]xcept in unusual categories of cases in which courts have developed no-duty rules, an actor's duty to exercise reasonable care does not require attention from the court."). Arguably, even the causal showing required for strict liability entails normative judgments to identify which counterfactual worlds are relevant to the causal question. *See id.* § 20 cmts. e-f ("[A] common response to the recommendation of strict liability for the causation of harm is that most accidents happen at the literal or figurative intersection of two or more activities. This is a reality that often makes factual causation indeterminate and insufficient as a strict-liability criterion."). For example, the assertion that "[w]hen the defendant, by blasting, projects debris that damages the plaintiff's property, common parlance might lead one to observe that that damage has been almost exclusively caused by the defendant's activity," only makes sense if we assume (i) the relevant question is what would happen if the defendant, as opposed to the plaintiff, had done something differently, and (ii) the other things we imagine the defendant doing, if not blasting, exclude other abnormally dangerous activities. *Id.*; *see also* Kenneth W. Simons, *The Restatement (Third) of Torts and Traditional Strict Liability: Robust Rationales, Slender Doctrines*, 44 WAKE FOREST L. REV. 1355, 1370-71 (2009) ("[A]ny notion that actor or activity *Y* has 'exclusively caused' harm to actor *Z* is unintelligible . . . [without] assumptions about . . . the legally relevant background entitlements . . . .").

[54] *See* Stapleton, *supra* note 40, at 448 ("[T]he conceptual framework and methodology of the Law provide filtering devices: that specify . . . relevant hypothetical comparator worlds.").

[55] *See, e.g.*, Charles A. Sullivan, *Tortifying Employment Discrimination*, 92 B.U. L. REV. 1431, 1436-37 (2012); Michael C. Harper, *The Causation Standard in Federal Employment Law: Gross v. FBL Financial Services, Inc., and the Unfulfilled Promise of the Civil Rights Act of 1991*, 58 BUFF. L. REV. 69, 69-70, 75-91 (2010); Brodin, *supra* note 48, at 313-18. Beyond academic assertions claiming as much, the Court is often explicit about drawing on tort doctrines. *See, e.g.*, Staub v. Proctor Hosp., 562 U.S. 411, 417 (2011) ("[W]e start from the premise that when Congress creates a federal tort it adopts the background of general tort law."). But, as we argue, the Court's general claims, absent more, do not determine which of the many possible formulations of the causal showing one endorses.

test . . . supplies the 'default' or 'background' rule against which Congress is normally presumed to have legislated . . . includ[ing] when it comes to federal antidiscrimination laws like § 1981."[56] Likewise, the Supreme Court has interpreted Title VII and other antidiscrimination statutes to include the "standard requirement of any tort claim," meaning causation in fact.[57]

However, the Court has been somewhere between vague and sloppy as to what the common law causation test is a test *of* when it applies this test to antidiscrimination claims. As we demonstrate in this section, the Court equivocates or conflates two different versions of this supposedly "ancient and simple" test—one that mirrors tort law by using a prior normative-legal concept (namely, *discrimination*) to narrow the relevant possible worlds to the counterfactual question, and another that bizarrely frames the counterfactual contrast in terms of the plaintiff's social status (e.g., as a man, as Black).[58]

But *which cause* must be shown to be the but-for cause of the plaintiff's alleged loss? Following the example of tort law, we think that the relevant causal test would ask the following counterfactual question using what we call the Normative Showing:

---

[56] *Comcast Corp.*, 140 S. Ct. at 1014 ("It is 'textbook tort law' that a plaintiff seeking redress for a defendant's legal wrong typically must prove but-for causation.") (quoting Univ. of Tex. Sw. Med. Ctr. v. Nassar, 570 U.S. 338, 347 (2013)).

[57] *Nassar*, 570 U.S. at 346.

[58] When interpreting antidiscrimination statutes, the Court switches between these two formulations. Sometimes the Court talks as if the cause needed to satisfy the but-for causal showing is that the defendant's "illegitimate" or "discriminatory" motive or act was the cause of the complained-of outcome. *E.g.*, *Comcast Corp*., 140 S. Ct. at 1014 ("Under this standard, a plaintiff must demonstrate that, but for the *defendant's unlawful conduct*, its alleged injury would not have occurred.") (emphasis added). At other times the Court talks as if the cause needed to satisfy the but-for causal showing is that the plaintiff's status was the cause of the complained-of outcome. *E.g.*, *Bostock*, 140 S. Ct. at 1753 (2020) ("Congress's key drafting choices—to focus on discrimination against individuals and not merely between groups and to hold employers liable *whenever sex is a but-for cause of the plaintiff's injuries*—virtually guaranteed that unexpected applications would emerge over time.") (emphasis added); City of Los Angeles, Dep't of Water & Power v. Manhart, 435 U.S. 702, 711 (1978) ("[T]he simple test [is] whether the evidence shows 'treatment of a person in a manner which *but for that person's sex* would be different.'") (emphasis added). Sometimes the Court will switch between the two within the span of a few sentences. *Compare* Price Waterhouse v. Hopkins, 490 U.S. 228, 241 (1989) ("It is difficult for us to imagine that, in the simple words 'because of,' Congress meant to obligate a plaintiff to identify the precise causal role played by *legitimate and illegitimate motivations* in the employment decision she challenges.") (emphasis added), *with Price Waterhouse*, 490 U.S. at 241-42 ("[I]nstead . . . Congress meant to obligate her to prove that the employer relied upon *sex-based considerations* in coming to its decision.") (emphasis added). *See also Price Waterhouse*, 490 U.S. at 237-38 ("[E]ven if a plaintiff shows that *her gender played a part in an employment decision*, it is still her burden to show that the decision would have been different if the employer had not discriminated.") (emphasis added).

> **Normative Showing:** If not for the defendant's *discriminatory* conduct, policy, motive, or intent, would the plaintiff have experienced this employment practice or loss?

Instead, as illustrated above in Part I, what we find is that the Court often states the counterfactual question using what we call the Status Showing:

> **Status Showing:** If the plaintiff had a different sex (or race, etc.), would they have suffered this employment practice or loss?

The slip between these two framings of the counterfactual question has rendered antidiscrimination causation doctrine deeply confused and backwards. A true transfer of the but-for test from tort law to antidiscrimination law would have us use the Normative Showing to understand the relevant counterfactual question. This framing would use a normative-legal concept of discrimination to delimit the relevant possible worlds, playing a role equivalent to duty in tort law: it requires us to answer what counts as discrimination *before* we can run the counterfactual causal test.[59]

In contrast, the Status Showing supposes that the question of whether unlawful discrimination occurred can be answered by, or simply reduced to, a question of causal dependence alone, as if the relevant but-for cause of discrimination is sex itself—a particular absurdity on any mere "biological" definition of sex.[60] Thus, rather than apply a counterfactual question framed using the normative-legal concept of discrimination, the Court appears to believe that a counterfactual causal test framed in terms of the plaintiffs' status will *deliver* a verdict as to what counts as discrimination. In doing so, they not only introduce

---

[59] *See supra* notes 10-15 and accompanying text and Part II-A.

[60] Both the Majority and Alito's Dissent assume that extension of the term "sex"—which, on their reading, must be the motive or cause of the plaintiff's complained-of outcome—is merely biological sex. *Compare Bostock*, 140 S. Ct. at 1739 ("[W]e proceed on the assumption that 'sex' signified [in Title VII] what the employers suggest, referring only to biological distinctions between male and female."), *with id.* at 1757 (Alito, J., dissenting) ("If 'sex' in Title VII means biologically male or female, then discrimination because of sex means discrimination because the person in question is biologically male or biologically female.") *and id.* at 1761 (Alito, J., dissenting) ("Title VII prohibits discrimination because of *sex itself*, not everything that is related to, based on, or defined with reference to, 'sex.'"). Yet, neither the Majority nor the Dissent explain what they imagine it means to have merely "biological sex" or "sex itself" as a cause or motive, (e.g., the person was fired merely because of their vagina or penis), as opposed to the social meaning, expectations, or stereotypes associated with sex (e.g., the person was fired because the employer believed working outside the home is not appropriate for persons with vaginas).

further indeterminacy into the causal test, they also set the causal test up for an impossible task.

We find the Court's construction of the test on clear display in *Comcast*. There, the Court decided that the but-for test requires that "a plaintiff bears the burden of showing that race was a but-for cause of [their] injury."[61] In other words, the *Comcast* Court explicitly doubles down on the idea that the cause needed to succeed on an antidiscrimination claim is the cause outlined by the Status Showing. The Court goes on to state, with respect to § 1981:

> The guarantee that each person is entitled to the "same right . . . as is enjoyed by white citizens" directs our attention to the counterfactual—*what would have happened if the plaintiff had been white*? This focus fits naturally with the ordinary rule that a plaintiff must prove but-for causation. If the defendant would have responded the same way to the plaintiff even if he had been white, an ordinary speaker of English would say that the plaintiff received the "same" legally protected right as a white person. Conversely, if the defendant would have responded differently but for the plaintiff's race, it follows that the plaintiff has not received the same right as a white person.[62]

By this point, the difference between the Status Showing and the Normative Showing should be clear. The Normative Showing requires the plaintiff to show that the defendant's discriminatory conduct, policy, motive, or intent was a but-for cause of their loss; the Status Showing requires the plaintiff to show that their protected status or a biological fact about their body (e.g., woman-qua-human with vagina) was a but-for cause of their loss. By using this Status Showing, the Court is clear that it is not imagining a close-as-possible world in which some discriminatory event is "completely and cleanly excised from history."[63] Rather, they are imagining (some) close-as-possible world in which the plaintiff has a different racial or sex status—that is, the counterfactual antecedent is one in which the plaintiff is (or is perceived to be) of a different sex or racial

---

[61] *Comcast Corp.*, 140 S. Ct. at 1014.
[62] *Id.* at 1015 (emphasis added). We disagree that the statutory language is properly, or even probably, interpreted as the Court suggests in the quoted text. Instead of calling for a counterfactual, the phrase the "same right . . . as is enjoyed by white citizens" could be read to articulate a substantive standard of rights to which other, *non*-White citizens are entitled.
[63] DAVID K. LEWIS, PHILOSOPHICAL LETTERS OF DAVID K. LEWIS: VOLUME 1: CAUSATION, MODALITY, ONTOLOGY 235 (Helen Beebee & A. R. J. Fisher, eds. 2020).

category—than the one they in fact are (or are perceived to be). *That,* we are told, will deliver a verdict as to whether discrimination because of sex or race occurred.

C. Problems with the Status Showing

We think that there are deep problems with relying on the Status Showing rather than the Normative Showing in the case of antidiscrimination law. One problem concerns the difficulty of conceptualizing social status categories like sex or race as "causes," and so what it means to alter these statuses in a counterfactual scenario. Within tort law, the relevant focus is typically on specific *events,* like driving 65 mph or producing bottles that are not childproof.[64] But this is not what we find within antidiscrimination law. Rather than focus on events, the Court instead wants to alter protected *statuses,* like being Black or being a woman.[65]

But what does it mean to change someone's status? For example, if we are imagining a female employee as a man, are we imagining that the employee has the same hair, voice, affect, upbringing, preferences, habits, etc. but different genitals? If we are imagining that a Black employee is white, are we imagining that the employee grew up in the same neighborhood and went to the same schools, except has a different set of ancestors?[66] For reasons we cannot understand, the Court seems to think that imagining counterfactual statuses is no more difficult than imagining counterfactual events. But readers familiar with causal inference in the social sciences, epidemiology, statistics, and computer science will know there are profound (to put it mildly) debates about this precise issue.[67] As we demonstrated in Part I, any time we engage with counterfactual

---

[64] *See* RESTATEMENT (THIRD) OF TORTS: LIAB. FOR PHYS. AND EMOT. HARM § 26 cmt. e (AM. LAW. INST. 2010) ("The requirement that the actor's tortious conduct be necessary for the harm to occur requires a counterfactual inquiry. One must ask what would have occurred if the actor had not engaged in the tortious conduct.").

[65] *See Comcast Corp.*, 140 S. Ct. at 1015.

[66] As discussed above, the Court seems happy to define sex as a matter of biology and to cash out sex counterfactuals as merely swapping genitals. *See* cases cited *supra* note 57 and accompanying text. But would they be as comfortable defining race as a matter of biology, and if so, what biological swaps would accomplish racial counterfactuals?

[67] Indeed, this precise question—whether social statuses such as race, gender, religion, etc. can be conceptualized as treatments (e.g. intervention events that can be administered to units without fundamentally changing the unit) in the counterfactual causal framework—has spawned a massive debate. *See generally, e.g.*, Paul W. Holland, *Statistics and Causal Inference*, 81 J. AM. STAT. ASSOC. 945 (1986); Donald B. Rubin, *Comment: Which Ifs Have Causal Answers*, 81 J. AM. STAT. ASSOC. 961 (1986); D. James Greiner & Donald B. Rubin, *Causal Effects of Perceived Immutable Characteristics*, 93 R. ECON. & STAT. 775 (2011); Jay S. Kaufman, *Epidemiologic Analysis of Racial/Ethnic Disparities: Some Fundamental Issues and a Cautionary Example*, 66 SOC. SCI. & MED. 1659 (2008); Issa Kohler-Hausmann, *Eddie Murphy and the Dangers of*

questions, we have to delimit the features of relevant possible worlds. It is extremely and notoriously difficult to say what this means when we are imagining changing a person's sex or race status.[68] And this, of course, is exactly the problem that led Gorsuch and Alito to opposite conclusions about the same question. While Gorsuch believes that "changing the plaintiff's sex" means holding fixed sexual attraction to the same sex (e.g., to men), Alito believes that it means holding fixed *same*-sex attractions.

To illustrate why causal dependence cannot tell us what counts as discrimination, consider another prohibited ground for employment discrimination listed in Title VII listed: religion.[69] If one thinks that the Status Showing can deliver an answer to the normative question of what counts as discrimination on the basis of a social status by "chang[ing] one thing at a time and see if the outcome changes," then it should work for all of the social statuses listed in Title VII.[70] We hope the following example illustrates that any claim that the dissimilar treatment of candidates from different social statuses with certain stipulated similarities is *discriminatory* rests on a prior normative premise—namely, that persons similar in those stipulated respects, but different in social status, are entitled to similar treatment.

Suppose an employer has a policy of prohibiting displays of religious observance in the workplace, meaning that religious employees cannot wear a symbol of significance to their particular religion. This employer then terminates a Christian employee for wearing a cross to work. Does this termination count as discrimination *because of religion*?

Under the *Bostock* Majority's reasoning, the answer would likely be "yes." Using that reasoning, "religious observance" would be conceptually separated

---

Counterfactual Causal Thinking About Detecting Racial Discrimination, 113 NW. U. L. REV. 1163 (2019) [hereinafter, Kohler-Hausmann, *Dangers of Counterfactual Causal Thinking*]; Lily Hu & Issa Kohler-Hausmann, *What's Sex Got to Do with Machine Learning?*, *in* FAT* '20: PROCS. 2020 CONF. ON FAIRNESS, ACCOUNTABILITY, & TRANSPARENCY, Jan. 27-30, 2020, at 513; Tyler J. VanderWeele & Whitney R. Robinson, *On the Causal Interpretation of Race in Regressions Adjusting for Confounding and Mediating Variables*, 25 EPIDEMIOLOGY 473 (2014).

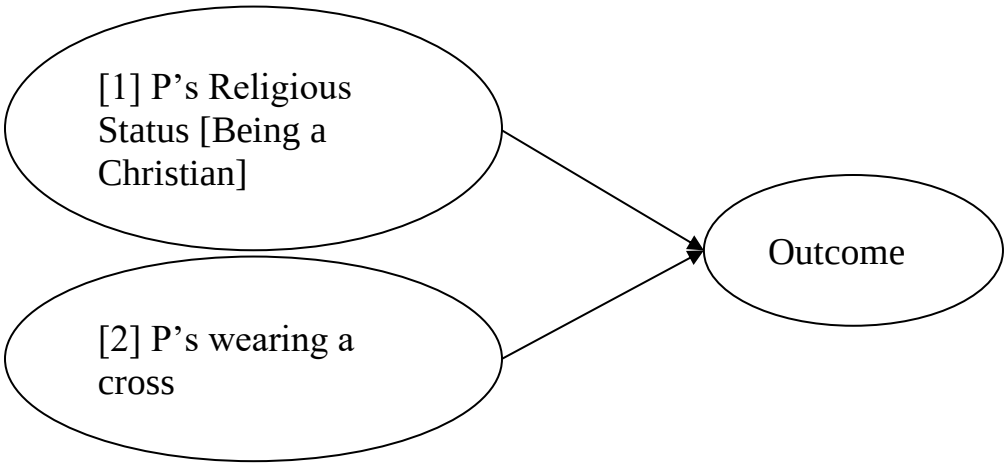
[68] [Citation TK]

[69] 42 U.S.C.A. § 2000e-2(a)(1) (West 2021).

[70] *See Bostock*, 140 S. Ct. at 1739 (2020) (discussing Title VII's but-for causation test generally while describing how it applies to "protected trait[s] *like* sex") (emphasis added). As a matter of doctrine, it seems that some members of the Supreme Court do not think that the but-for test for disparate treatment discrimination is the same for religion as it is for the other statuses protected by 42 U.S.C. § 2000e-2(a)(1). *See, e.g.*, E.E.O.C. v. Abercrombie & Fitch Stores, Inc., 575 U.S. 768, 773-74 (2015) (arguing that an employer whose motive is to "avoid[] accommodation" might violate Title VII even where they have "no more than an unsubstantiated suspicion" that some religious accommodation may be required, and even if the employer would take the same action if an employee of a different religion or no religion at all asked for the same accommodation).

into two parts: [1] the plaintiff's religion [Christian] and [2] the relevant symbol or behavior.[71] We capture this reasoning in Figure 3 (an analogue to Figure 1):

Figure 3:



With that separation, we could reason that an employee with a different religion—say, Islam—but the same symbol or behavior, would *not* be fired because the cross is not a symbol of religious observance for Muslims.[72] Therefore, firing the Christian employee would be deemed religious discrimination if we hold fixed the symbol or behavior (e.g., [2]) while toggling the employee's religion (e.g., [1]). As Gorsuch similarly reasoned in *Bostock*, this reasoning would suggest that religion was a "but-for" cause of the outcome.

By contrast, if we use Alito's reasoning, the answer would likely be "no." Why? Mirroring Alito's argument and using his strikethrough text gimmick, [73] we would need to add two additional employees to the mix: a Muslim who wears a headscarf to work, and a Christian who wears a headscarf to work. We now have four example employees, with the fired employees crossed out below:

---

[71] See *supra* fig. 1 and notes 26-27 and accompanying text.

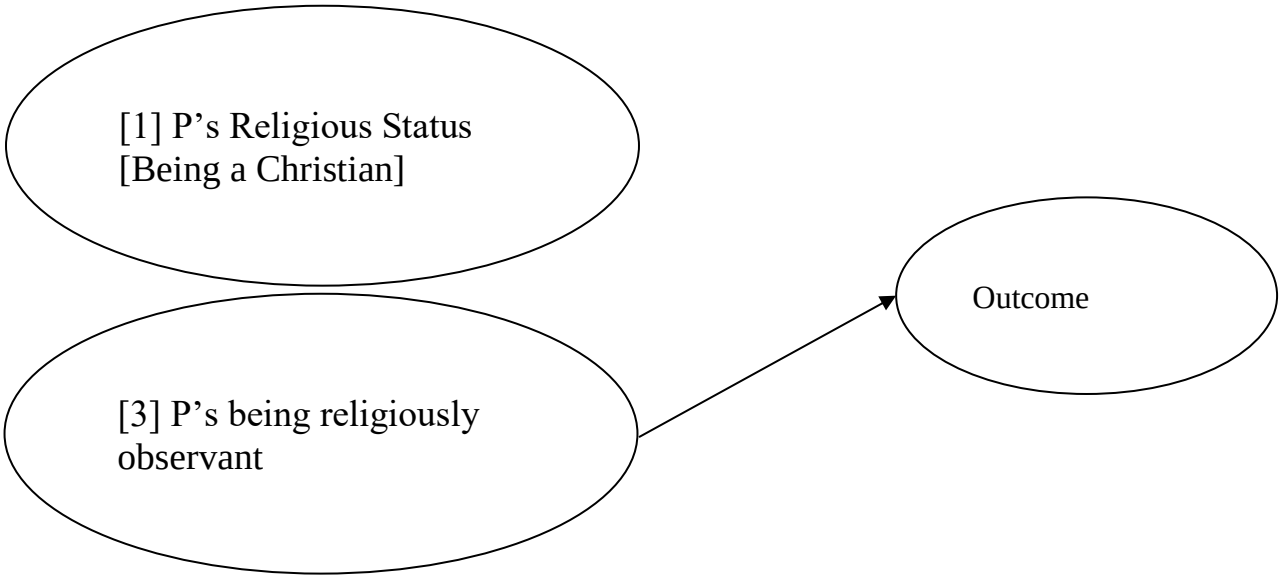
[72] Analogously, for a person taken to have female reproductive features assigned at birth, a skirt is not a symbol of gender "non-conformity," nor is a male sexual partner a symbol of—to use Alito's term—"homosexual orientation." *Bostock*, 140 S. Ct. at 1763 (Alito, J., dissenting).

[73] *Id.* (Alito, J., dissenting) ("We now have the four exemplars listed below, with the discharged employees crossed out:
~~Man attracted to men~~
Woman attracted to men
~~Woman attracted to women~~
Man attracted to women.").

~~Christian + cross~~
Muslim + cross
~~Muslim + headscarf~~
Christian + headscarf

According to Alito, the fired employees "have one thing in common."[74] It is not being Christian, being Muslim, wearing a cross, or wearing a headscarf. It is being *religiously observant*. So, using Alito's approach, we could conclude that the employer is not guilty of discrimination on the basis of *religion*, but on the basis of *religious observance*. We capture this reasoning with an analogue to Figure Two:

Figure 4:



Our point here is not that Gorsuch's use of the counterfactual test is better than Alito's. Again, as we argued at length in Part I these status counterfactuals are inherently indeterminate.[75] Rather, we are making the point that the Status Showing does not, by itself, tell us whether changing a person's religion (or sex, race, etc.) in a counterfactual scenario and observing a different outcome means

[74] *Id.* (Alito, J., dissenting) ("The discharged employees have one thing in common. It is not biological sex, attraction to men, or attraction to women. It is attraction to members of their own sex—in a word, sexual orientation. And that, we can infer, is the employer's real motive.").
[75] *See supra* Part I.

the act/practice was discriminatory. We cannot determine whether religious discrimination occurred by just looking to how the employer treats a Muslim who wears a cross, or a Christian who wears a headscarf. The question of religious employment discrimination is what rights or freedoms people are owed in the workplace given specific religious meanings and practices. Crosses, for example, bear a particular significance within Christianity. The claim of freedom of religious expression might require, for a Christian employee, the freedom to wear a cross in the workplace. We cannot know if a particular Christian employee experienced religious discrimination by looking at how crosses are treated in general, or by noting that Muslim employees are also forbidden from wearing headscarves. We can *only* know this by examining the *social reality of religious practice*, and thinking about what we, as a society, want to affirm as the limits of freedom of religious expression in the workplace.

Return to Alito's dissent in *Bostock* and its analogy in the religious discrimination hypothetical. We hope you share our view that banning *everyone* from religious observance in the workplace does not answer whether banning workplace religious observance is discrimination on the basis of religion because to determine whether such ban is discriminatory we still need to answer the normative question of what kind restrictions on religious practices we, as a society, are willing to tolerate. As we have repeatedly tried to demonstrate, cause in fact cannot answer a normative question.[76] The counterfactuals in the religion example only tell us exactly what we already knew: that the employer is opposed to employees wearing symbols of religious observance.[77] Similarly, Alito's counterfactuals regarding a gay employee only tell us exactly what we already knew: that the employers in *Bostock* and *Zarda* are opposed to employees believed to have same-sex attractions.

Alito argues in his dissent that firing all gay employees would exempt an employer from the charge of sex discrimination because it shows that the employer's true target is sexual orientation, not sex per se. But this is absurd in the same way that showing that an employer's true target is religious observance, not religious classification per se, exempts an employer from the charge of religious discrimination. One can, for the sake of argument, grant that the causal

[76] *See supra* notes 45-53 and accompanying text.

[77] They also show us that one can't even come up with candidates for the counterfactual thought experiment without content-specific knowledge of the social category of *religion* or *gender*. How else do we know if the person is being, for example, "religiously observant" when they don't wear a cross or a headscarf (as opposed to another random fashion choice)? Similarly, without content-specific knowledge, how do we know if a person is being "gender conforming" or "gender /non-conforming" with respect to their sexual attractions and presentation?

contrasts expressed in the strikethrough text examples are true, and even that the employer's intent was to pick out employees who stand in a particular relation to their assigned sex or religion (e.g. those that stand in a "culturally nonconforming way" or an "observant relation"). But we are still left with the normative question of whether acts or intents based on an employee's standing in *that particular relation* to the relevant sex or religious status are discriminatory.

Rather than follow tort law and outline its counterfactual question using the Normative Showing, the Court over and over again assumes that mere causal relations will *deliver* a normative result—specifically, a determination of whether or not unlawful discrimination occurred. Rather than take discrimination to be an event that may or may not have caused a bad outcome, the Court treats the plaintiff's social status as an event that, if a but-for cause of some bad outcome, renders the defendant's conduct discriminatory.[78] In tort law, but-for causation is an independent element of a tort claim. It is not a way to prove that the legal-normative grounds for imposing liability on the defendant—breach of duty—was present.[79] Yet, in the context of antidiscrimination law, the Court recasts causation as an evidentiary route to proving that the legal-normative grounds for imposing liability on the defendant—discrimination—was present.

Causation describes a (metaphysical) fact about the world, that two things stand in a relation of counterfactual dependence. It does not evaluate whether it is good or bad for that dependence to hold. We need a normative theory to answer a normative question, and the normative question at issue in antidiscrimination cases is whether employers are entitled to disadvantage employees for standing in *a particular relation* to the relevant racial, sex, religious, etc. statuses. Mere causation will never give us this normative theory.[80] Simply because an employer who fires gay men would *also* fire gay women does not mean that employers are not obligated to give employees freedom of sexual and gender expression. Even if we concede that gay men and gay women share the "commonality" Justice Alito proposes, nothing follows from it. His argument is non-sequitur and gets us no closer to determining how men and women *ought* to be treated with respect to their sexual and gender expression. What we need is a prior theory of what discrimination is. Only once we have such a theory and have applied it, can we ask what we think is the relevant question in these cases: Were it not for the

---

[78] *See Comcast*, 140 S. Ct. at 1015 (2020) ("To prove a violation, then, the government had to show that the defendant's challenged actions were taken '"on account of"' or '"by reason of"' race—terms we have often held indicate a but-for causation requirement.'") (quoting Gross v. FBL Financial Services, Inc., 557 U.S. 167, 176-77 (2009)).

[79] *See* Coleman *supra* note 51, at 439.

[80] *See supra* notes 45-53 and accompanying text.

defendant's discriminatory action or policy, would the plaintiff have suffered a loss?

D. Interpretive Aside

Although we promised that this essay is not about the "right" interpretation of the law according to some school of statutory construction, at this point you might object that, even if the Normative Showing is better than the Status Showing, the text of Title VII calls for the Status Showing. Specifically, you might argue that the text 42 U.S.C. § 2000e-2(a)(1) does not prohibit a type of normative conduct left undefined in the text (i.e., discrimination). Rather, you might argue, that the statute prohibits a class of conduct that is defined in the text: namely, any act caused by a particular motive or mental state—i.e., sex, race, religion, etc.[81]

You might propose that this interpretation is mandated by two provisions. First, it follows from reading the statutory text prohibiting "discriminat[ion] against any individual . . . because of such individual's race, color, religion, sex, or national origin" to mean that the employer's motive regarding an employee's sex, race, etc. caused the action.[82] Second, you might propose it follows from the text of the 1991 amendment to Title VII, which says, in relevant part, that "an unlawful employment practice is established when the complaining party demonstrates that race, color, religion, sex, or national origin was a motivating factor for any employment practice, even though other factors also motivated the practice."[83] Specifically, you might propose to interpret 42 U.S.C. § 2000e-2(m)

---

[81] *See, e.g.*, Berman, Mitchell N. and Krishnamurthi, Guha, Bostock was Bogus: Textualism, Pluralism, and Title VII (February 1, 2021) at 27-34. Notre Dame Law Review, Forthcoming, Available at SSRN: https://ssrn.com/abstract=3777519 or http://dx.doi.org/10.2139/ssrn.3777519; Eidelson, Benjamin, Dimensional Disparate Treatment (September 1, 2021). Harvard Public Law Working Paper No. 21-25, Southern California Law Review, Vol. 95, No. 4, 2022, Available at SSRN: https://ssrn.com/abstract=3915787 or http://dx.doi.org/10.2139/ssrn.3915787; ANDREW M. KOPPELMAN, BOSTOCK AND TEXTUALISM: A RESPONSE TO BERMAN AND KRISHNAMURTHI 9-11 (Northwestern Univ. Pritzker Sch. of Law, Pub. L. and Legal Theory Series No. 21-11 ed. 2021) ("The statute [Title VII] is concerned with motives, not causes. So it is motivation that a court is looking for when it asks 'whether the evidence shows "treatment of a person in a manner which but for that person's sex would be different."'") (quoting 42 U.S.C. § 2000e-2(m)). However, even as described by Koppelman, it is not clear if he, and authors making similarly arguments, read the statute to forbid employment actions animated by certain motives—meaning beliefs or reasons that explain why a willed act was taken—or actions animated by mental states—meaning something akin to purpose, intent, or knowledge about an attendant circumstance of the act, such as whether a specific employee has a particular race, color, religion, sex, or national origin status.

[82] 42 U.S.C. § 2000e-2(a)(1).

[83] 42 U.S.C. § 2000e-2(m).

to *define* the prohibited conduct referenced 42 U.S.C. § 2000e-2(a)(1)—i.e., the normative concept of discrimination—as *any* employment practice where an individual's "race, color, religion, sex, or national origin was a motivating factor."[84]

What, then, counts as a sex motive or reason with respect to sex per se? Following both the Majority and Alito's Dissent in *Bostock*, we might define it as a motive or reason with respect to the employee's status as "biologically male or female."[85] There are two things to note about this position. First, it defines any employment practices of sex segregation or classification—including use of gendered pronouns (e.g., 'he' and 'she') and the practices dissenting Justices in *Bostock* vehemently want to defend as non-discriminatory, such as sex-specific"*[b]athrooms, locker rooms, [and other things] of [that] kind*," as discrimination because sex is a motive of or reason (read: but-for cause) for the employment practice.[86]

Second, interpreting the statutory text to prohibit employment practices caused by the motive or reason of "*sex itself*,"[87] means a plaintiff must be under 2000e-2(m)'s 'mixed-motive' provision anytime sex per se and some other factor

---

[84] If you read of Title VII to prohibit any act/practice where the protected statuses listed under 42 U.S.C. § 2000e-2(a)(1) was the motive, it would seem to have radical implications for at least one status listed under Title VII: religion. For example, if an employer gives a scheduling accommodation to an employee to observe religious services they would, in theory, be committing religious discrimination because accommodating that individual employee's specific religion was the employer's motive or purpose in taking the employment action. This would lead to a strange tension between the main prohibition of Title VII under § 2000e-2(a), and § 2000e(j), which defines "religion" to include "all aspects of religious observance and practice, as well as belief, unless an employer demonstrates that he is unable to reasonably accommodate …." 42 U.S.C. § 2000e(j). On your reading, the main section § 2000e-2(a) would prohibit discrimination qua taking religion as a motive for action, but the definitions section § 2000e(j) would then compel discrimination, but only for one of the statuses (religion) protected under § 2000e-2(a). Defining failure to accommodate religious practices as discrimination under all circumstances just means you are not applying the same motive-based test for one status in the statute. In at least one case, some members of the Supreme Court proposed that failure to reasonably accommodate religious practices was identical to disparate treatment, while others argued that failing to accommodate a religious practice is not intentional discrimination unless the religious nature of the practices was part of the reason or motivation. *Compare Abercrombie & Fitch Stores, Inc.*, 575 U.S. at 772 n.2 (2015) (majority opinion) ("[F]ailure to hire 'because of' the plaintiff's 'religious practice' [here, wearing a head scarf] . . . is *synonymous* with refusing to accommodate the religious practice. To accuse the employer of the one is to accuse him of the other."), *with Abercrombie & Fitch Stores, Inc.*, 575 U.S. at 783-84 (Thomas, J., dissenting in part) (arguing that defining intentional discrimination to include actions taken because of an employee's conduct [here, wearing a head scarf] "that *happens* to be religious" would ultimately penalize employers who acted without any such discriminatory motive).

[85] *Bostock*, 140 S. Ct. at 1739, 1757 (2020).

[86] *Id.* at 1778 (Alito, J., dissenting).

[87] *Id.* at 1761 (Alito, J., dissenting).

were both reasons or motives for the employment practice. Take an example like *Price Waterhouse*, where the plaintiff claimed that she was disfavored for (among other things) not wearing makeup and being aggressive at work.[88] If you insist that discrimination consists solely in having sex qua "biologically male or female"[89] as a reason or motive for the action, then this plaintiff presents a mixed motive case. A motive regarding wearing makeup or being aggressive at work is not identical to one regarding reproductive organs per se, and both of the former two motives were also necessary to bring about the employment action. Even a paragon case of "No Women Need Apply" policy involves mixed motives about the employee's biological sex and about what roles are proper for persons with that believed sex. You would be hard pressed to come up with a single case where biological sex is the singular sufficient but-for cause of the employment action.

Furthermore, the view that a motive of "*sex itself*"[90] is equal to discriminatory motive would imply that almost every plaintiff is properly classified as proceeding under 2000e-2(m)'s mixed motive provision, and thereby limited to declaratory relief. Recall that in mixed motive cases, plaintiffs are not eligible for damage awards or reinstatement injunctive relief under the other portion of the 1991 amendment, 42 U.S.C.A. § 2000e-5(g)(2)(B).[91]

Further, interpreting 42 U.S.C. § 2000e-2(m) to define the conduct referenced 42 U.S.C. § 2000e-2(a)(1) is not how most courts, including the Supreme Court, understand the mixed motive provisions. The Court has not read § 2000e-2(m) to define what is referenced 42 U.S.C.A. § 2000e-2(a)(1), but rather has construed it to provide that liability is established under Title VII (albeit with limited remedies) where the adverse employment outcome had multiple but-for causes, so long as one such cause was discrimination.[92] As the Supreme Court recognizes, the 1991 amendments were drafted, in part, to overrule that portion of the *Price Waterhouse* decision that allowed a defendant-employer to escape a finding of liability altogether if it could show that it would have made the same employment decision in the absence of discrimination.[93] Congress decided to

---

[88] *Price Waterhouse*, 490 U.S. at 235 (1989).
[89] *Bostock*, 140 S. Ct. at 1757 (Alito, J., dissenting).
[90] *Bostock*, 140 S. Ct. at 1761 (Alito, J., dissenting).
[91] 42 U.S.C. § 2000e-5(g)(2)(B)(i)-(ii) ("(B) On a claim in which an individual proves a violation under section 2000e-2(m) of this title . . . the court . . . shall not award damages or issue an order requiring any admission, reinstatement, hiring, promotion, or payment, described in subparagraph (A).").
[92] *See Nassar*, 570 U.S. at 343, 355 (2013) ("For one thing, § 2000e–2(m) is not itself a substantive bar on discrimination. Rather, it is a rule that establishes the causation standard for proving a violation defined elsewhere in Title VII.").
[93] *See* Desert Palace, Inc. v. Costa, 539 U.S. 90, 94-95 (2003) ("[Section] 107 of the 1991 Act . . . 'respond[ed]' to *Price Waterhouse* by 'setting forth standards applicable in "mixed motive" cases'

provide that a plaintiff that was in fact the subject of discrimination—but would have suffered the employment action even without it—is still entitled to a declaration of wrongdoing, though not to damages or injunctive relief such as reinstatement.[94] The Supreme Court seems to think a plaintiff is under the mixed motive provision of 42 U.S.C. § 2000e-2(m)—and therefore, subject to the remedy restrictions of § 2000e-5(g)(2)(B)(i)-(ii)—only when discrimination (as opposed to sex) was mixed with a non-discriminatory/lawful cause.[95] This interpretation, which understands the things that are mixed in so-called mixed motive cases as discriminatory and non- discriminatory motives or causes (as opposed to sex and non-sex motives or causes) is consistent with the Normative Showing of the but-for causation requirement. Thus, the § 2000e-2(m) provision does not define "discrimination" in § 2000e-2(a)(1); we still need to know *under what conditions* having an individual's race, color, religion, sex, or national origin as a motive for an act counts as a discriminatory motive.

III. TOWARD AN ALTERNATIVE VIEW

The question of discrimination is an inherently normative one—one that asks us what is a wrongful way to treat someone on the basis of a protected status, be it a religious status, a sex status, or some other protected status.[96] The religious observance example discussed above shows us that answering whether an action is discriminatory requires a rich social understanding of these statuses, as well as a normative theory of what people deserve or are otherwise owed *in light of* what statuses mean in our society. Only by first answering whether something is discriminatory—and thereby apply the Normative Showing—will we arrive at a helpful framing for administering the but-for causal test, because our answer will delimit what an employer *ought* to have done. From here, we can then ask: If the

---

in two new statutory provisions."); William N. Eskridge Jr., *Title VII's Statutory History and the Sex Discrimination Argument for LGBT Workplace Protections*, 127 YALE L. J. 322, 375 (2017) ("In committee reports [about the 1991 amendments], the sponsors were clear that the section aimed at *Hopkins* only 'overrules one aspect of the decision.' One committee report emphasized that these amendments would in no way affect *Hopkins*'s holding that 'evidence of sex stereotyping is sufficient to prove gender discrimination.'") (internal citations omitted).

[94] *See* 42 U.S.C. § 2000e-5(g)(2)(B)(i)-(ii).

[95] For example, in *Desert Palace*, the Court considered whether "a plaintiff must present direct evidence of discrimination in order to obtain a mixed-motive instruction," and the Court held the answer is no. *Desert Palace*, 539 U.S. at 98. The reason this case was even treated as a "mixed motive" was because the defendant maintained that they had non-*discriminatory* reasons for the firing (specifically, that the plaintiff had a long history of volatile relations with co-workers).

[96] *See supra* Parts II-A and II-B.

employer had done as they ought to have done, would the plaintiff have suffered this loss?

To answer the question of what counts as discrimination, we believe that the Court must rejoin this normative question of *what people are owed* to the question of what occurs "because of sex." No act or policy regarding sexuality or gender expression is discriminatory in a vacuum of its relationship to what sex means in our society, just as a policy prohibiting crosses and headscarves is not discriminatory in a vacuum of its relationship to what those symbols meant to specific religions and what religion means in our society. Only with a theory of what is wrongful in light of the social meanings and inequalities attending sex status will we arrive at an answer as to whether firing gay or transgender employees is discriminatory because of sex. Arriving at such a theory is no trivial matter. The categories addressed by antidiscrimination statutes are attended by a host of complex and contested differences. But our point is that the very concept of discrimination "on the basis of" any given status cannot be unearthed simply by noting that *other* people with the same status were or would be treated similarly, or that people without that status were or would be treated differently. It can only be identified once we know what people are owed, notwithstanding—or in some cases because of—the differences that attend these social statuses.

To motivate this framing shift, let's first consider the following hypothetical case: An employer refuses to hire someone, citing the fact that they dress in an androgynous fashion. When that person files a complaint, the employer responds, "Their sex has nothing to do with it! I would fire any employee, male or female, who was confusing to look at."

We take it for granted that sex discrimination occurs this case.[97] The Majority's reasoning in the *Bostock* case, when applied to this case, would struggle to reach this conclusion because "androgyny" is not defined in terms of violating the norms associated with a *specific* sex status.[98] No doubt, Alito would argue that "androgynous" is gender-neutral, just as he claims that being "gay" or "transgender" is gender-neutral, which would in turn further demonstrate his

[97] If you disagree, no doubt you can come up with a different, analogous case, such as where the employee is bisexual, not married, has no children, or anything else where the status/behavior described at low level of granularity is not a sex-specific stereotype (the way that "having sex with men" is a female-specific stereotype), but nonetheless—is a sex-related stereotype that functions to ground the way sex functions as a stratifying social system.

[98] That is, there is no obvious way to set up the causal diagram in Figure 1 so that if you "hold constant" androgynous clothing (even if described at a lower-level of generality like "pants and plain tee shirt") and toggle sex of the plaintiff, you will get a contrast in outcomes.

unclear use of the term "gender-neutral."[99] While we would not cast doubt on the Court's ability to finagle their reasoning to suit this outcome, there is a simpler, logically cleaner way to the decision that sex discrimination occurred.

Rather than separate "discrimination" from "because of sex," we propose that the best reasoning is to consider whether the outcome is one where the employee was treated poorly relative to how he or she *ought* to be treated, *given* background facts about social inequalities, stereotypes, norms, and social meanings surrounding their sex status. We think that a policy that prohibits employees from dressing in an androgynous fashion, whether male or female, is wrongful in light of the social meanings of sex statuses, as any such policy reinforces nefarious social norms according to which there is distinct ways that men and women *ought* to dress—norms that maintain the illusion of exacerbated sex differences between men and women and assign material and dignitary worth on the grounds of conformity to those norms.[100] The Majority's mistake is to interpret the prohibition expressed by the phrase "discrimination because of sex" as picking out a class of things that are discriminatory because they "caused" by sex, instead of picking out a class of things that are discriminatory because they reinforce or act on nefarious social meanings of sex statuses.

By adopting the Normative Showing, we can arrive at a reasonable logic for antidiscrimination statutes. This much is clear from common-sense reasoning about action that is alleged to be wrongful *because of* some group-based status. The claim that an individual's encounter in a particular instance is discrimination *because of* a group-based status is a statement that, "the distinctive wrongfulness of the action or practice is dependent upon what the category [ ] consists of."[101] An individual's claim that some act counted as discrimination *because of* their group-based status means that, notwithstanding the relevant norms, practices, assumptions, or associations that attach to that group, the person is *owed* some sort of right, access, or consideration. But we can't answer that question unless we explicitly state what sorts of norms, practices, assumptions, or associations that attach to *that group*, and then explicitly stake out what we think is fair or just in various domains in light of those norms, practices, assumptions, or associations.

---

[99] A similar point could be raised about another case: Suppose an employer fires a female employee, citing the fact that she is too emotional in the workplace. When the employee files a complaint, the employer responds, "The fact that she is female has nothing to do with it! I would fire any employee, male or female, who was acting like a little girl at work."

[100] MARILYN FRYE, THE POLITICS OF REALITY: ESSAYS IN FEMINIST THEORY 17, 28 (1983) ("This matter of our sexes must be very profound indeed if it must, on pain of shame and ostracism, be covered up and must, on pain of shame and ostracism, be boldly advertised by every means and medium one can devise.").

[101] Kohler-Hausmann, *Dangers of Counterfactual Causal Thinking*, *supra* note 67, at 1179.

We hope this view will also explain why the refrain "Title VII protects individuals not groups" has never made much sense to us.[102] In so far as such language only means that something can be discriminatory without categorically harming every person ascribed membership to the group at issue, that seems uncontestable. Even a rule like "No women need apply" does not categorically exclude or harm every woman. Some women may have no interest in applying to that job, or even think it is immoral or wrongful for women to work outside of the home. For them, the rule "No women need apply" does not count as a disabling exclusion or even a subjective wrong because it does not conflict with what they believe women should be doing. But this is a paradigm instance of sex discrimination, because the rule is based on and reinforces nefarious ideas about what ought to be open to persons irrespective of sex stereotypes and traditional roles.

The refrain "Title VII protects individuals not groups" is thus misleading because Title VII and related statutes require us to have some theory of what counts as "discrimination *because of*" whatever group-based statuses are named in the statute, e.g. sex, religion, or race, and to distinguish such discriminatory conduct from other kinds of conduct that is legal even if objectionable for other reasons. The statutory language that makes it "unlawful [ ] for an employer to fail or refuse to hire or to discharge any individual, or otherwise to discriminate against any individual . . . because of such individual's race, color, religion, sex, or national origin," [103] addresses the rights workers are owed in their raced, sexed, religious, statuses, not as individual abstracted from these statuses. The statute does not forbid employers from using many forms of categorizations for making distinctions, exclusions, or qualifications.[104] Employers are even allowed to indulge grounds for distinctions that are expressly irrational, random, or even mean spirited.[105] Therefore, the very terms of the statute demand that we

---

[102] *See Bostock*, 140 S. Ct. at 1745 ("[T]he statute focuses on discrimination against individuals, not groups.").
[103] 42 U.S.C. § 2000e-2(a)(1); *see also Price Waterhouse*, 490 U.S. at 239 (1989) ("In passing Title VII, Congress made the simple but momentous announcement that sex, race, religion, and national origin are not relevant to the selection, evaluation, or compensation of employees.").
[104] *Price Waterhouse*, 490 U.S. at 239 ("Yet, the statute does not purport to limit the other qualities and characteristics that employers *may* take into account in making employment decisions. The converse, therefore, of 'for cause' legislation, Title VII eliminates certain bases for distinguishing among employees while otherwise preserving employers' freedom of choice.")
[105] *Bostock*, 140 S. Ct. at 1762 (Alito, J., dissenting) ("[O]ther than prohibiting discrimination on any of five specified grounds . . . Title VII allows employers to decide . . . [even] idiosyncratic criteria are permitted; if an employer thinks that Scorpios make bad employees, the employer can refuse to hire Scorpios. Such a policy would be unfair and foolish, but . . . it is permitted.") (citation omitted).

formulate some way of identifying when an individual has suffered discrimination *because of* (for example) sex or race, and distinguish it from distinctions that are not on the basis of a protected status.

Legal readers may, at this point in the argument, worry that our positive proposal applies only in cases of disparate *impact* claims, but not disparate *treatment* claims. Under standard doctrinal views, disparate impact claims address intents, act, or policies that, while they do not target a protected category *per se,* produce a pattern of outcomes for persons of a protected category that we (through various doctrinal tests) deem intolerable.[106] In contrast, disparate treatment claims are understood to be claims that address intents, acts, or policies that target a protected category *per se* —for example, in the case of sex, a policy of refusing to hire women.[107] Those who accept this distinction might be tempted to argue that only disparate impact claims must, logically, be defended by pointing to *why* the intent, act, or policy in question is wrongful *in light of* the meaning of the protected form of categorization, but not disparate treatment/intentional discrimination. The latter, on this view, counts as discrimination only on the grounds that the intent, act, or policy in question makes an explicit distinction on the basis of (for example) sex.

While we cannot fully argue for this view here, this objection fails because *every* claim of unlawful discrimination must rest on normative judgments about what persons are owed, given social meanings attached to the relevant status (sex in our running example). The doctrinal division between disparate impact and disparate treatment relies on the notion that disparate treatment identifies a class of intents, acts, or policies that count as "discrimination because of sex" due to nothing more than the fact that they target, e.g., sex or race *per se*. But that is an illusion.

Consider again the paradigmatic example of exclusion noted above: "No women need apply." Let's assume, as the Majority does in *Bostock*, that the intent, act, or policy is meant to differentially treat (and specifically, to exclude) every woman. Even this is not sufficient to show that this policy is *wrongful* or *discriminatory because of sex*, because—as should be clear to conservatives on the Court who insist that differential treatment with respect to pronoun use, accessing bathrooms or locker rooms does not count as discrimination—there may be in instances where differential treatment of people based on social status

[106] *Griggs v. Duke Power Co.*, 401 U.S. 424, 436 (1971).
[107] *Washington v. Davis*, 426 U.S. 229, 239 (1976).

is desirable (even if it is not in this example).[108] In each case, we still must explain *why* this policy is wrongful, given the nefarious stereotypes, expectations, and norms of treatment that attend being a woman in our society.

All of us (we hope) can come up with many reasons to justify the moral conclusion that excluding women from labor force opportunities is wrongful in light of the historical and current meanings of sex difference and stratification. And perhaps because that proposition is so widely accepted, the normative reasoning fades from view. But this is an essential part of the logic in *both* so-called disparate impact and disparate treatment claims. And that is precisely our point: an evaluative judgment that discrimination based on a protected category occurred (e.g., sex) *always* relies on normative reasoning about what treatment is *wrongful in light of that category*. Whether an act or policy targets a protected category directly (e.g., "no women allowed") or indirectly (e.g., "no people under 5'8" allowed") does not change this underlying reasoning.

Now return to *Bostock.* To answer whether firing gay employees is discrimination because of sex, the Justices attempted to answer whether the outcome (firing) was caused by sex.[109] We have shown that this inquiry does not answer whether the act counts as *discrimination because of sex*. We should, instead, ask whether being fired for being gay is wrongful in light of sex-based social norms that police the sexuality of all persons. Social norms that say that men ought only to have sex with women (and women with men), are nefarious norms that are rooted in sexism. We think people should not lose their jobs for deviating from these sex-based norms, and that regardless of the employee's sex, firing someone for not conforming to these social expectations is wrongful—it is discrimination because of sex. But the question of whether employers ought to be permitted to police their employees according to these sex-based norms is precisely the question that we must answer to make a determination of discrimination. Causation alone will not get us there.

---

[108] The provision of bona fide occupational qualification ("BFOQ") for sex under Title VII—regardless of whether one endorses their specific exemptions—shows that the evaluative judgment that an act or practice is discrimination because it does not logically follow from the fact that it consists in a distinction on the basis of sex *alone*, without recourse to a normative account for why that distinction is wrongful in light of the category of sex. It is the latter sorts of judgments that make many of us conclude that many of the proffered BFOQs are in fact discrimination because of sex, such as "being a woman dressed in a sexy uniform is a BFOQ for the job of casino cocktail server." Ann C. McGinley, *Babes and Beefcake: Exclusive Hiring Arrangements and Sexy Dress Codes*, 14 DUKE J. GENDER L. & POL'Y 257, 260 (2007).

[109] *See supra* Part I.

IV. A CAUTIONARY CONCLUSION

The Supreme Court has a supremely confusing way of talking about causation in antidiscrimination law. Their approach, we've shown, hides normative judgements concerning what is discrimination "because of" a group-based status like sex beneath a counterfactual question about what would have occurred but-for the plaintiff's status. In other words, the Court smuggles in normative evaluation of an employer's actions or policies under the guise of facts about causal inference. Our approach lays bare the normative reasoning that—as a conceptual matter—necessarily lies at the heart of any inquiry into what counts as discrimination because of a social status, such as sex or race. Moreover, it is a much more logical interpretation of the statutory text: something is discriminatory "because of sex" when what makes it wrongful—in the way that Title VII is concerned with *wrongful* conduct—is something about sex social statuses.

Of course, our claims suggest that judges must give up the myth that hard cases can be decided with value-neutral methodology. But when Congress passed statutes with broad normative terms like "discrimination," they explicitly delegated to courts the difficult task of designating acts or practices in the world as instances of a normative kind. We think that those with different substantive views about what is owed should welcome open and honest debate about what kinds of differential treatment is wrongful in light of the social meanings of sex statuses. Similarly, and in direct contrast to the machinations of the *Comcast* Court, we think that determinations of race discrimination must examine what kinds of differential treatment is wrongful in light of racial social meanings and stratification.[110]

Despite the Supreme Court's supreme confusion about causality and normativity in antidiscrimination statutes, we are hopeful there is still an opportunity for the legal scholarly community and litigators to move the doctrine in a more coherent direction. We offer these reflections to start such a dialogue.

---

[110] Defining discrimination in terms of but-for causation is equally concerning in the context of race discrimination. The *Comcast* Court limited itself to declaring that the "ancient and simple 'but for' common law causation test [ ] supplies the 'default' or 'background' rule against which Congress is normally presumed to have legislated when creating its own new causes of action," including antidiscrimination statutes. *See Comcast*, 140 S. Ct. at 1014 (2020). Accordingly, there is less to say about how the Supreme Court applied this so-called "test" to the particular facts of that case. Nonetheless, what the Court announced is nothing short of a nonsensical formula that gives license to disguise overly restrictive concepts of discrimination beneath appeals to causation. Our argument is not that the but-for "test," as used for discrimination, is a stingy test, or that it does not reach what we think it ought to reach. Rather, we argue it is an incoherent appeal to turn causal relations into normativity. It currently is, and will continue to be, used to obfuscate objectionable views about what is discriminatory in light of the meaning of race in our society.