Political Philosophy

Hu, Lily. 2025. Sex Discrimination, Normativity, and Begging the Causal Question. *Political Philosophy*, 2(1):262–289.DOI:https://doi.org/10.16995/pp.19782

Open Library of Humanities

Sex Discrimination, Normativity, and Begging the Causal Question

Lily Hu, Philosophy, Yale University, US, lily.hu@yale.edu

Most leading philosophical accounts of discrimination theorize discrimination as a *causal* notion: roughly, an action discriminates on the basis of X (e.g., race, sex) if X *causes* (in the right way) the adverse outcome. This article explores the prospects for this causal account. Focusing my attention on the case of sex discrimination, I argue that struggles to settle the causal question regress to question-begging. This argument calls into question not just whether the causal account can handle sophisticated cases, which have recently been conceived of (by some) as discriminatory on the basis of sex, but even classic cases of sex discrimination. The negative argument in turn brings to light an alternative *normative-causal* approach to discrimination, which I sketch and suggest is promising indeed.

Sex Discrimination, Normativity, and Begging the Causal Question

Lily Hu
Philosophy, Yale University, US

Recent developments are pressing a shift in philosophical interest in discrimination. From legal challenges that push on the conceptual bounds of sex and race discrimination to growing worries about the threat of algorithmic discrimination, the question of *what it is to discriminate on the basis of sex or race, in the first place* now emerges as a surprising puzzle that turns out to have been left largely unaddressed in the extant legal and philosophical literatures on the subject.[1] What exactly constitutes discrimination on the basis of sex or race or religion or any other "protected" grounds? What does it take to discriminate on these bases?

The most influential proposals so far suggest a *causal* account: roughly, an action discriminates on the basis of X (e.g., race, sex) if X *causes* (in the right way) the adverse outcome. Slight variations on this theme lurk behind several leading philosophical theories of discrimination, where causation figures as an explication of the connection that stands between the grounds of discrimination, X, and the discriminatory action—often referred to as the discrimination concept's "because of" or "on the basis of" condition.[2] In the United States, courts have recently solidified their commitment to

[1] I should here note two recent works on discrimination that also draw attention to the literature's broad tendency to overlook the question. Benjamin Eidelson (2022, p. 785) writes that scholars and judges alike have "failed to provide [] a coherent, general account of what it means for an action to be taken 'because of' an attribute in the sense relevant to claims of disparate treatment." Deborah Hellman (2023, pp. 206–207) frames the matter as a "definitional" issue—"Despite the centrality of the concept of disparate treatment over many years, we are lately coming to recognize that we don't quite know what it is"—that, in her eyes, sets "a research agenda for our times."

[2] Of the four most prominent philosophical monographs on discrimination published in the past 15 years or so, all three which discuss the concept of discrimination explicitly endorse some version of the causal account. Causation figures centrally in: Benjamin Eidelson's (2015, pp. 17–24) account as the "explanatory condition" of discrimination, in Kasper Lippert-Rasmussen's (2013, pp. 36–40) discussion of the "because of" condition in his section on "Because," and in Sophia Moreau's (2020b, pp. 19–21) discussion of what unifies direct and indirect discrimination. Hellman's (2008) book is the only one that does not discuss causation, since it focuses rather exclusively on the wrongmaking feature of discrimination and so does not consider the conceptual question.

causation as an essential component of a discrimination claim in the landmark Supreme Court case *Bostock v. Clayton County*, which ruled that discrimination on the basis of sexual orientation and transgender status are instances of discrimination on the basis of sex.[3] The Court's decision hung entirely on its but-for analysis of causation: If the adverse outcome would not have obtained were it not for the plaintiff's sex, then we have the kind of sex causation that is hallmark of sex discrimination.[4]

Besides aligning with prevailing US legal doctrine, the causal account promises to deliver distinctively philosophical boons, too. An appeal to causation presents a compelling solution to a persistent question in discrimination theory. Most philosophical work takes there to be two distinct strands of discrimination: disparate treatment (or direct) and disparate impact (or indirect). What unifies them? The causal account answers that both are cases where some protected status *causes* some disadvantage. Discriminatory "intent," "nonaccommodation," and the broadly felt distributive harms characteristic of disparate impact all mark different ways that this causal link may manifest.[5] What's more, the causal solution does not just clarify this conceptual puzzle but has been taken to secure a normative upshot that many theorists of discrimination have sought to vindicate: a more expansive notion of discrimination that unsettles the primacy of disparate treatment within it. Add to this other advantages deriving from outside of philosophy and law—for instance, the fact that causal analyses are dominant among social scientific approaches to empirically "detecting" discrimination[6]—and the causal account of discrimination appears not just well-supported but neat and cohesive.

These formidable credentials notwithstanding, there is a notable gap in the story. Despite several leading discrimination theorists' explicitly marking out the important

[3] *Bostock v. Clayton County*, 140 S. Ct. 1731 (2020).

[4] What matters for the purposes of this article is *Bostock*'s strident affirmation of the conceptual connection between discrimination and causation rather than the substantive content of its causal analysis.

[5] Moreau (2020b, pp. 20–21) writes that "we do not need to interpret 'on the basis of' as referring to a causal chain part of which always runs through the discriminator's mind. For the causal chain might instead extend from the rule or practice to the discriminatee and the group that shares the relevant trait with her. This alternative way of thinking of the phrase 'on the basis of' gives us a unified interpretation of direct and indirect discrimination." For a discussion specific to US discrimination law, see Zatz 2017. Now, whether unification of these different strands of discrimination *should* be a desideratum of a theory of discrimination is a separate matter, which I do not address in this article. Here I simply take for granted that most theorists of discrimination seem to take it as such.

[6] See e.g., Greiner and Rubin 2011.

role that causation plays in their accounts, the causal analysis that is supposed to do this crucial theoretical work has itself remained opaque. This is especially true of the two paradigmatic cases of discrimination, discrimination on the basis of race and on the basis of sex, where longstanding controversies over what race and sex *are* make analyses of race and sex causation particularly vexing. How exactly are the accounts of race and sex causation which are so central to accounts of race and sex discrimination supposed to work? Can tracing the causal effects of race or sex really explain why certain cases are or are not discriminatory on these bases, as has been widely presumed?[7] The aim of this article is to explore the first question and provide a partial answer to the second.

In what follows, I will argue that a causal analysis of discrimination is beset by a serious worry of circularity. To fix ideas, I focus my attention on the case of sex discrimination and show that the dialectic that emerges out of struggles to settle the causal question regresses to question-begging. Causal analysis cannot establish or explain whether an action is discriminatory on the basis of sex, because figuring whether sex causes a given outcome raises the question of what effects *confound* rather than *constitute* sex's causal influence. And, as I will argue, to answer this is just to answer the central question at issue.[8]

Here is how the argument is laid out. I begin in Section I by recounting the causal approach to theorizing the discrimination concept. I will not here endorse any particular causal account but instead home in on some minimal features that all analyses that define discrimination as an essentially difference-making causal concept must contain. In Section II, I introduce the circularity challenge and then show in Section III how the argument extends beyond a narrow set of cases and calls into question whether the causal account may handle even classic cases of sex discrimination. My conclusion is that attempts to escape the circle will not work to rehabilitate the causal account. This negative result brings to light an alternative positive proposal for a causal account of discrimination, one which grants normative judgments about sex discrimination

[7] For an illuminating discussion that traces where race (or a perception of race) figures in the causal chain that leads up to a discriminatory action, see Singh and Wodak 2024. For a philosophy of science angle on the same issue, see Weinberger 2022.

[8] Although this article's argument is formulated with respect to sex, it should become clearer as it unfolds that many of the complexities that lurk behind the question of what defines sex's causal effects carry over to the question of what defines race's causal effects. A full elaboration of this version of the argument and a parallel challenge to a causal account of racial discrimination, however, calls for future work.

a place in a theory of sex causation. I sketch the outlines of such a *normative-causal* approach to discrimination in Section IV and close in Section V with remarks on why I think such a route is promising indeed.

I. A CAUSAL ACCOUNT OF SEX DISCRIMINATION

Philosophical analyses of the concept of (non-moralized) discrimination typically feature three successive elements. First, discrimination is an essentially comparative notion. At issue in a discrimination claim is the fact that one person is treated one way, whereas "another person"—I use quotations to denote that this person may not be actual—is, or would be in some appropriate hypothetical circumstance, treated another way. Hence, *differential* treatment, actual or hypothetical, is a necessary feature of discrimination.[9] Second, to qualify as discrimination, this differential treatment must be keyed to some feature, which is the basis or *grounds* of discrimination. Differential treatment with no basis whatsoever, i.e., entirely arbitrary differential treatment, cannot be discrimination. The third element describes what this *keying* amounts to, that is, what the link must be between the grounds of differential treatment and the treatment itself for it to constitute discrimination on those grounds.[10]

Philosophers of discrimination are in broad agreement about how to characterize the first two elements, but the third is a point of active scholarly debate. One way of filling it in is to claim that discrimination on the basis of X requires that X be the agent's *reason* for her differential treatment. Another way claims that it suffices that X merely *motivates* her differential treatment for her action to be discriminatory on the basis of X. Yet another claims that the differential treatment is keyed to X in the right way if X *causes* (in the right way) the different treatment.[11] I'll call this last way of filling in the above analysis a *causal account of discrimination*. By and large, advocates of the

[9] This should be distinguished from the claim that *what makes discrimination wrong* is an essentially comparative matter. This article does not focus on the normative basis for the wrongness of discrimination but rather on a core feature of the so-called "non-moralized" concept itself: that for an act to even qualify as discrimination, wrongful or not, it must involve differential treatment of some kind. Such views take it that, as Lippert-Rasmussen (2013, pp. 16) puts it, "discrimination is *essentially comparative*."

[10] See fn. 2.

[11] Technically, X itself need not be what causes the action. It could simply be that the discriminating agent has some, true or false, belief about X or perceives, veridically or mistakenly, the discriminatee to have X. For a discussion of "misperception" cases of discrimination, see Singh and Wodak 2024.

causal account adopt a difference-making analysis of causation: that causes make a difference to whether their effect obtains; so no cause, no effect. On this view, what it is for X to cause the differential treatment at issue—and thus for the treatment to be discriminatory on the basis of X—is for it to be the case that, were it that −X, that treatment would not obtain.

The causal difference-making idea is thought to easily explain why cases of exclusion due to overt interpersonal racism and sexism are discriminatory. But its advantage over alternative analyses really lies in its ability to make sense of those cases of discrimination that manifest in more subtle, non-one-to-one, non-explicit expressions of animus. In particular, hostile workplace dynamics, longstanding social practices with differential group impacts, and implicit cognitive biases are now broadly taken to be potential mechanisms of discrimination. Yet many theories struggle to explain why. The causal account dispenses with this difficulty. Conscious or unconscious, interpersonal or structural, what unifies these cases is the presence of a causal connection between an individual's belonging to (or perceived belonging to) the relevant social category and the disadvantage they suffer at the hands of the workplace culture, the social practice, their employer's biases, and so on. The following cases are illustrative. Granting for the sake of argument that Sexual Harassment, Appearance Guidelines, and Implicit Bias are bona fide cases of sex discrimination, the causal account successfully explains why.

Sexual Harassment

> Gloria's place of work has a culture of sexual harassment against women. She refuses her harasser-coworkers' and -managers' sexual advances. Were Gloria not perceived to be sexed female, she would not be so targeted, and would not be fired for refusing her boss's advances.

Appearance Guidelines

> It is standard practice at Sasha's place of work for women to wear a dress and heels. She does not do so. Her employer cites her violation of the company's appearance guidelines as the reason for her being let go. Were Sasha not a woman, she would not be in violation of the appearance guidelines and would not have been let go.

Implicit Bias

> Ameera's responses in her performance review were perceived by her manager as *brusque* and *bossy*. The manager was not consciously aware of the role that Ameera's being a woman played in his evaluation of her performance. He would

> not have evaluated her so poorly and denied her the promotion were it not for her being a woman.

Recent discussion of the causal account of discrimination, however, has revolved not around these cases but another, trickier set of cases, which proponents of the causal account also claim to be able to explain. I will here group them together as SEX+ CASES.

> SEX+ CASES
>
> Alex is a transgender girl who is prohibited from using the girls' bathroom at school. Were Alex not assigned "male" at birth, she would have been free to use the girls' bathroom.
>
> Bill, a man, is fired from his work upon his boss's finding out that his partner Andrew is a man. Were Bill a woman and married to Andrew, "she" would not have been fired.

SEX+ CASES will play a central role in the remainder of this article—but not as a "test case" for whether the causal account can or cannot deliver the "right" verdicts about these cases. Rather, I will be setting to the side whether cases of discrimination on the basis of transgender status and sexual orientation are or are not cases of sex discrimination, in order to ask whether a causal account of discrimination can deliver conclusory verdicts on these cases *at all*—be they affirmative or negative. I will argue that it cannot. SEX+ CASES thus serves as my entry point into diagnosing a set of issues that afflict the causal account of discrimination on the basis of sex more broadly. Working through how the account struggles to deliver causal verdicts and thereby discrimination verdicts in these cases will preview the difficulties it faces in explaining other, far less controversial and supposedly far easier to analyze, cases of sex discrimination. These will be the cases in my argument that really endanger the viability of a causal account of discrimination.

Two final notes before proceeding. First, this article focuses on the normative phenomenon of "sex discrimination" or "discrimination on the basis of sex"; I take these to be terms of art that for various historical and legal reasons use the term "sex" rather than the term "gender," even though the latter may seem to readers today to be more apt. The matter of what the conceptual scope of "sex" in "sex discrimination" is will be a central question in this paper. I do not mean that question to be settled or even considerably informed by the use of the word "sex" in this terminology.

Second, in the wake of *Bostock*, many scholars have offered their own critiques of SEX+ CASES. But thus far, these arguments reject SEX+ CASES on the grounds that these *particular* causal analyses are flawed, incomplete, or draw problematically on undefended normative choices. I do not wish to deny these claims. But I take these responses to issue only a partial dissent to the causal account of discrimination. They do not argue against causal analyses of discrimination in toto, only that the specific analyses offered in Sex+ Cases fail.[12] My aim in the next two sections is to show why *all* such causal accounts of what constitutes sex discrimination necessarily fail to resolve these cases. This will require attending more closely to the inner-workings of causal analyses than previous discussions have thus far, and focusing, in particular, on the question of what it is for sex to be a cause.

II. BEGGING THE CAUSAL QUESTION

Sex+ Cases brings out the difficulty of carrying out a difference-making analysis that bears witness to the causal relevance of *sex* to the outcomes in question, rather than other factors that may confound the inquiry at hand. So while proponents of the causal analyses in Sex+ Cases use the cases to show that the treatment in question is indeed

[12] Mitchell Berman and Guha Krishnamurthi (2021), for example, reject the causal analyses in Sex+ Cases in favor of the alternative, still causal analyses in Sex+ Cases* (below). I consider and respond to their account in Section III. Robin Dembroff, Issa Kohler-Hausmann, and Elise Sugarman (2020, p. 7) present arguments that are closest to challenging the causal account of discrimination in toto. But their conclusion that a causal account is unworkable is, in my view, drawn too quickly from the mere fact that there are multiple causal contrasts to which a causal account might appeal. For example, their remark that a given causal analysis is "not an independent test of discrimination; it is an expression of one's normative priors about what is discriminatory" can be read as conflating the reasons for adopting a view and the reasons for the view itself. It might well be true as a psychological matter that what motivates particular individuals to choose one counterfactual contrast over another are their normative priors about discrimination. But this does not preclude there still being good grounds for choosing one contrast over another. In a different paper, Dembroff and Kohler-Hausmann (2022, pp. 67–68) summarize their conclusions about the failure of the causal account as follows: "once we take 'because of sex' to mean 'sex as a but-for cause,' there is nothing in... the nature of causation that requires any particular view as to which counterfactual possibilities are the relevant ones to consider." The reason, they claim, is that given the context of the causal query, the only way of deciding among possible worlds is to pick the one that best fits the "normative and legal definition of unlawful discrimination because of sex." But they do not explain why this should be the only way of settling the question of possible worlds, and they do not present arguments against alternative approaches to the question (such as Berman and Krishnamurthi's for instance). All this is to say that although I ultimately agree with Dembroff and Kohler-Hausmann in their conclusion that the proposed causal accounts fail and in their gestures towards why they fail, their explanations are incomplete. My arguments in Sections II and III will show why.

caused by sex, their opponents retort that the comparators used to generate these causal verdicts suffer from causal confounding. The charge is this: Changing Alex's and Bill's sex status makes changes to other sex-related facts about them: whether they are transgender or cisgender, and whether they are gay or straight. To allow these factors to vary across comparators is to allow their causal influence to be "picked up" as the causal relevance of sex and as a result, to count as discriminatory on the basis of sex. But whether that is so—whether the causal operation of these factors *are* instances of the causal operation of sex—is precisely what the causal analysis is supposed to test *for*. The test just *is* a test of whether treatment caused by gender identity and sexual orientation is treatment caused by sex. Thus, to set up the casual analysis in this way is to rig the game by presupposing the answer that the analysis is itself meant to prove.[13]

What is more, these opponents claim, alternative ways of running the analyses suggest that the adverse outcomes suffered by Alex and Bill are positively *not* caused by their sex status. According to the analyses in Sex+ Cases*, causation *does not* vindicate these cases as instances of sex discrimination.

Sex+ Cases*

> Alex is a transgender girl who is prohibited from using the girls' bathroom at school. Were Alex assigned "female" at birth and still transgender and seeking use of "his" preferred bathroom at school, "he" would be similarly prohibited.
>
> Bill, a gay man, is fired from his work upon his boss finding out that his partner Andrew is a man. Were Bill a woman and gay, "she" would have still been fired.

But does Sex+ Cases* really manage to avoid begging the question? Its comparators are constructed by changing sex and holding fixed Alex's being transgender and Bill's being gay. But what justifies holding these factors fixed? Holding them fixed is only methodologically licensed, after all, if they might *confound* sex's causal effects. Otherwise, doing so will yield comparators that do not fully vary in the target variable

[13] I follow the convention of discussing the causal analysis in terms of being a "test" for discrimination. This might suggest a purely epistemic reading of the analysis: that the purpose of the causal "test" is to help us figure out whether discrimination occurred. That is not how I read the causal analysis. As I see it, the causal "test" is no mere epistemic device for determining whether an act was discriminatory. Rather, proponents of a causal account of discrimination take the causal connection to be a necessary condition for discrimination. Thus, that the treatment at issue is in fact caused by, say, sex is a part of what it *is* for an act to be discriminatory on the basis of sex. This in turn means that the circularity that I claim besets extant causal accounts of discrimination is not merely epistemic but in fact metaphysical.

of interest, which will in turn generate the wrong causal conclusions. To presume, then, that these factors indeed risk confounding is to presume that their effects are *distinct* from sex's and so must be struck out of any observed differences across the comparators. But now we are back at the question precisely at issue: *whether* these effects—effects of gender identity and sexual orientation—*are* effects of sex? So, Sex+ Cases* begs the question just as Sex+ Cases does.

In sum, a causal account of sex discrimination determines whether an action constitutes discrimination on the basis of sex by analyzing sex's causal relevance to the treatment in question. Doing so requires setting up causal contrasts that may tell the causal role that sex plays. But in the cases at hand, candidate causal factors are conceptually related to one another. As a result, there are multiple ways of setting up a causal contrast in which the target variable "sex" is different—with different ways of setting up the contrast bringing in different "other" changes into the analysis (e.g., sexual orientation, trans/cis gender status). The trouble is that settling on any particular contrast as the right one requires settling on that which is precisely in question: whether an effect of these "*other*" changes is or is not an effect of *sex*. So, the causal account cannot *establish* whether an action is discriminatory on the basis of sex, because *defining* what it is for sex to have a causal effect raises the question of what effects confound rather than constitute sex's causal influence (in order to set up the right causal contrast). And to answer this is just to answer the central question at issue.

Note the contrast between this argument and the typical objections that scholars discussing *Bostock* have put forth. Those objections tend to revolve around the mere fact of there being multiple changes brought in tow by the change to sex status in both Sex+ Cases and Sex+ Cases* and the need therefore to find grounds on which to privilege one over the other.[14] This framing sets up the problem as one of tie-breaking between two legitimate ways of determining sex's status as a cause for the purposes of determining sex discrimination. By contrast, my contention is that each way of setting up the causal contrast begs the central question at issue: whether some factor's influence constitutes or confounds the effect of sex. Thus, the source of the problem I draw out is *internal* to the causal analysis, rather than deriving *externally* from a need to out-compete another candidate analysis. The internal critique suggests that the problem with the causal account's handling of these cases has deeper roots than prevailing explanations have noted.

[14] See fn. 12. The multiple changes that appear in the causal contrasts in Sex+ Cases* are these: When Alex is assigned female at birth and is still transgender, Alex now looks to use the boys' rather than girls' bathroom. Bill, as a gay woman, is no longer attracted to men, and so is no longer partnered to Andrew.

In the next section, I will move to show how these worries about circularity indeed apply more widely than just the cases in Sex+ Cases. We will arrive at that broader challenge by way of first considering some possible lines of response to the argument above. This will serve to further elucidate how the argument works and lay the groundwork for seeing how it will extend.

III. EXTENDING THE CHALLENGE

My main claim thus far has been that the causal account cannot explain or establish whether the actions at issue in Sex+ Cases are or are not cases of sex discrimination, because running the causal analysis requires drawing on an account of what defines the causal effects of sex. Is an effect of a factor conceptually tied to sex—such as sexual orientation or trans/cis gender status—an effect of sex? Or distinct from it? Alas, this is precisely the question at hand: whether differential treatment caused by these factors is differential treatment caused by sex.

That argument may have seemed fast. So this section will further elaborate on it, by considering two attempts to sidestep the metaphysical circularity I diagnose above. These look to ground the choice between Sex+ Cases and Sex+ Cases* on some prior independent account that distinguishes sex and its causal effects from sexual orientation and gender identity and their causal effects (or shows that there is no way of distinguishing these). I consider two such strategies. One looks to ground the choice of causal comparator in a *metaphysical account* of the category sex and its causal behavior. The second looks to do so by considering what constitutes a *mental representation* of sex rather than a non-sex or merely sex-related representation.

A. The Metaphysical Category Approach

The metaphysical category approach looks for independent metaphysical grounds on which to control for (or not control for) gender identity and sexual orientation in figuring the causal effect of sex. This strategy draws on a metaphysical account of the category sex and its causal profile that distinguishes it (or does not distinguish it) from that of these latter categories. For example, such an account might hold that the causal effects of sex status are essentially effects of biological facts, whereas the effects of gender identity and sexual orientation are essentially effects of social facts. This demarcation would then justify controlling for the causal effects of these social factors across contrasts in order to identify the causal effects of a set of biological factors.

The problem with this way of proceeding is that the paradigmatic discriminatory effects of sex in the social world are triggered not by biological factors but by social factors. So, an approach that varies a narrow set of biological facts about a person and controls for all the markedly social facts about them—facts about, say, their gender presentation—cannot make sense of even classic cases of sex discrimination. Take, for example, the canonical case of employment discrimination where an employer refuses to consider any woman who applies for a job opening. The employer refuses an application by a candidate named "Brittany" because he assumes that it is submitted by a woman. The employer is presumably not averse to, say, uteruses or certain chromosomal makeups as such. Rather, he is sensitive to a *social* fact: someone with the distinct social status of being sexed female, not someone who meets some set of biological criteria.[15] After all, if "Brittany" had a different biological sex status—if, say, "Brittany" were sexed male or intersex—but the manager still perceived them to be a woman, their file would presumably be discarded all the same. But then sex, on this view, really did not cause the employer to reject the application, since changing the relevant set of biological facts about a person and controlling for all the social facts, such as their name and perceived social status of being a woman, does not lead to a different outcome. So, the causal account of discrimination paired with a biological account of sex cannot identify this clear-cut case of sex discrimination.[16]

The biological account is meant here only as an illustrative example of how the metaphysical category approach to the causal analysis works: by defining and *delimiting* the scope of sex and its causal profile to mark it as distinct from nearby social categories. But seeing how the account fails imparts a lesson that applies to the strategy more broadly. Paradigmatic cases of sex discrimination involve apprehension

[15] Nor is it the case that "Brittany's" status of being a women function as a mere "proxy" for what the manager is truly concerned with: whether she has some set of sex organs.

[16] What if the employer's perception of "Brittany" as a woman is itself caused by a perception of them as, say, sexed female? Then wouldn't sex "indirectly cause" the application to be discarded? It first bears marking that this causal structure presumes that these two perceptions are distinct. This is a substantive assumption, which I discuss in greater detail in Section III.B. But even granting it, there remains the question of whether the employer's perception of "Brittany" as a woman should or should not be held fixed in figuring the causal effect of his perception of "Brittany" as sexed female. To assume that it must not be held fixed is to commit, by transitivity, to counting its causal effects as sex's causal effects. This begs the question. On the other hand, if it is held fixed, then since the employer discards the file all the same, the case is not one of discrimination on the basis of sex. I return to this case and analysis in Section III.B and discuss an analogous dilemma to it in a version of Implicit Bias in Section III.C.

of a distinctively *social* fact: for instance, that of occupying the position of being a woman and all that tends to come along with this status in a society structured by gender. This means that although an approach to causal analysis, which draws on a metaphysical account of sex based in biological facts may successfully ground a causal contrast on non-question-begging grounds, it does so at the cost of being able to explain paradigmatic cases of sex discrimination, which are caused by a distinctively social conception of sex. In other words, commitment to a metaphysics of sex that sharply distinguishes it from all nearby social categories related to sex *does* provide non-circular independent grounds for setting up causal contrasts for determining the causal role of sex. The problem, however, is that the resulting causal analysis cannot make sense of even the most basic cases of sex discrimination. Surely this is too steep a price to pay. If the causal account is to have any claim to being a plausible theory of sex discrimination, it must get these cases correctly.

This leaves the theorist taking this approach to adopt a metaphysics on which sex acts as a cause in ways that overlap with how some social markers of sex act as causes such that these social markers should not be controlled for in the proper causal contrast. But now, she is back at square one. Which social markers exhibit causal behaviors that are constitutive of, rather than distinct from, sex's causal behaviors? This is precisely the question at hand. Does sex act as a cause via its nearby social markers, gender identity and sexual orientation? Or are the effects of these categories distinct from sex's?

This might now seem precisely where the difficult projects of doing social metaphysics or social science about sex begins. But despite the fruitfulness of these areas of theorizing writ large, I am skeptical of their ability to overcome the specific circularity problems that arise here for two reasons. First, a causal analysis of sex discrimination must be able to determine the causal effect that sex has in a given setting. This involves attending to certain behaviors and outcomes and sorting which differences in those behaviors or outcomes are due to sex and which are not. Prevailing accounts of sex, however, are too coarse-grained to be able to carry out this exercise for all circumstances where the discrimination question might arise, without the help of other resources. Thus, even granting for the sake of argument broad consensus on these issues of social metaphysics or social science, I am doubtful that this best general theory of sex will not need to be supplemented by answers to such questions of constitution and confounding and suffice on its own to deliver determinate verdicts of sex discrimination in all cases. Second and more importantly, much debate about the best social-metaphysical and social-scientific account of sex revolves around what sex's causal effects are. Consider, for an analogous example, arguments about the metaphysics of race that draw on various phenomena of racism and racial discrimination to support a realist and social constructivist account of the category.[17] Here we see causal theorizing and disagreements about race and sex's causal effects *driving* disagreement about social-metaphysical and social-scientific accounts of race and sex. As such, it is hard to see how one can rely on consensus about the latter to resolve debate about the former. And so long as a general theory of sex cannot resolve all such questions, we are left with only circular reasons for choosing Sex+ Cases or Sex+ Cases*.

B. The Mental Representation Approach

The mental representation approach ends up needing to thread a similarly fine needle, although from a rather different starting place. This approach considers not the categories themselves but the discriminator's *mental representations* of these categories. On this view, if an agent's representation of a person's sexual orientation causally affects their behavior in a distinct manner than does their representation of the person's sex, then the right comparator is one in which while the "sex" mental representation should vary, the "sexual orientation" mental representation should be controlled for.

Mitchell Berman and Guha Krishnamurthi seem to endorse the representation approach to filling in the causal account of discrimination on the grounds that what the relevant comparator is for a given causal inquiry is a contextual matter, and the context of a causal account of discrimination is a "motivational analysis," which attends to how things appear "to the employer's mind."[18] The right choice of causal contrast, according to Berman and Krishnamurthi, is the one that "does 'least violence' to the correct description and explanation of the actual motivating reasons of the actor."

Here I will grant for the sake of argument Berman and Krishnamurthi's contextualist approach to the causal analysis as well as their claim that the discrimination context dictates that the causal analysis centers on an agent's mind and motivations. Still, what *is* the "correct description" of these mental states? This is precisely the crux of the matter for the representation approach. We need an account of such mental representations of a person *as sexed female* or *as transgender* or *as gay.* Eidelson, for instance, writes that an agent who discriminates on the basis of P "represents a person as falling under the description" P, "regards" them "P-wise," or has a "P-perception" of them.[19] But

[17] See e.g.: Àsta 2018; Haslanger 2019.

[18] Berman and Krishnamurthi (2021, pp. 106, 113) write that the task for a causal account of discrimination is "to pick out the closest worlds in the context of a motivational analysis." This way of resolving the indeterminacy is cited favorably in Eidelson 2022, p. 796.

[19] Eidelson 2015, p. 17, 21.

how are these mental states characterized exactly? Only if an agent's "sex-proper"-perception may be delineated from their "sexual orientation"-perception," only if the "sex-proper" representational content can be separated out from the "trans/cis gender status" representational content, will this approach be able to determine what must be controlled for (and held fixed) and what must vary in the right causal contrast.

At this juncture, there are two options for how to fill in the content of these representations. The first retreats entirely into the agent's subjective account of their motivations and thereby conflates the important distinction that Eidelson notes between the perceptions and mental representations themselves and the agent's view of these.[20] This tack would thus seem to strike causation out of the analysis entirely. For if what must be controlled for in the right causal contrast is the agent's motivations as the agent themselves delineates them and specifies their contents, then the analysis ceases to be a genuinely causal one at all. What results is an account that gives an entirely motivational account of discrimination: what is to discriminate on the basis of sex is for the agent to take themselves to have been motivated by sex in acting. The *causal* analysis becomes mere window dressing.

The second option is to distinguish these different sex-related representations by drawing on an objective metaphysical account of these mental states and their different causal profiles. But this lands us back with the very same set of issues that we encountered above with the metaphysical category approach. The route that seeks a hardline distinction between the causal profile of a "sex-proper" mental state, on the one hand, and a "sex-related social factor" mental state, on the other, fails to yield an extensionally adequate theory of sex discrimination. To see how, return to our earlier canonical case of employment discrimination. The employer refuses to consider any application from anyone he takes to be a woman. He receives an application from someone called "Brittany" and immediately discards it. Did the employer's sex-representation of "Brittany" cause this outcome? Consider a view that cleanly delineates a representation of someone as sexed female from a representation of someone as a woman. The reasoning, for instance, might be that the former representation is of someone falling under a biological description, to borrow Eidelson's language, and the latter is of someone falling under a social description. Now suppose the employer represents "Brittany" both as sexed female and as a woman. Then, on this view, the causal analysis would hold fixed the "woman"-representation in order to home in on the causal effect of the "sexed female"-representation. And since the employer declines the application all the same so long as he takes "Brittany" to be a woman,

[20] This distinction is very important in Eidelson's theory, for it is what separates his account of (direct) discrimination from those that center an agent's intentions.

it is *not* the "sex-proper"-representation then that is doing the causal work. So, the causal account fails to vindicate what is by presumption a genuine canonical case of sex discrimination.[21]

The natural response, of course, is to say that the representation of the candidate as a woman *counts as* a representation of them as sexed and so should not be held fixed in the causal analysis. But, just as we saw in the preceding section, to grant this move is to open the door to distinctively *social* representations—such as that of a candidate's being a woman—being "sex-proper" representations. But opening the door to *some* such "sex-related social factor" representations as "sex-proper" representations raises the question of *which* "sex-related social factor" representations exhibit causal behaviors that are a part of, rather than distinct from, the causal behavior of the "sex-proper" representation—which returns us to the core question precisely at issue.

C. Drawing Lessons Beyond Sex+ Cases

The preceding two subsections identify a critical stumbling block for a causal account of sex discrimination. Any causal analysis of sex may proceed only by delineating the causal factor that is "sex-proper" from the factor that is (merely) "sex-related." This is prerequisite to determining what causal controls must be in place in the properly conducted causal analysis. Call this the *causal delineation* problem.

The causal delineation task must navigate two distinct challenges. On the one hand, the distinction must be able to fit with canonical cases of sex discrimination. An analysis of sex causation that fails to account for any such cases of sex discrimination simply does not give an analysis of sex discrimination. On the other hand, it must not beg the question at hand. The question just is whether discrimination on the basis of a *sex-related factor* (e.g., sexual orientation, gender identity) counts as discrimination on the basis of *sex*. And the causal analysis must itself be what provides the answer.

I have shown that the causal theorist cannot successfully navigate this dilemma in the cases at issue in Sex+ Cases. In those cases, no causal analysis may deliver a verdict without begging the question; the causal account thus falls on the second horn. In this section, I move to show that this argument has wide ramifications. For while the difficulty of non-circularly grounding the right causal contrast to determine sex's causal status comes to the fore very clearly in those cases, parallel problems emerge in the account's handling of "simpler" cases, too.

[21] Needless to say, the employer who has no sex-representation of "Brittany" at all cannot be discriminating on the basis of sex in declining the application (on this view).

Let's start by returning to Sexual Harassment. Why is the causal account (presumed to be) able to rule the case a case of sex discrimination? Gloria is fired when she resists her boss's sexual advances. The intuitive causal analysis centered on the following comparison case: Were she sexed male, the employer would not have solicited sex from "him" and so would not have fired "him" for resisting. Sex causes the firing.

But now with the causal delineation problem in view, we might question whether that really is the right comparator. After all, why shouldn't the *employer's sexual attraction* be held fixed across the comparators? It seems to me rather plausible that sexual attractiveness and sex should be considered factors sufficiently distinct from each other such that the former *should* be controlled for in determining the causal relevance of the latter. Berman and Krishnamurthi's analysis would appear to suggest the same—lest the causal analysis do "'violence' to the correct description and explanation of the actual motivating reasons of the actor": the employer's sexual attraction. But on this way of carrying out the causal analysis, sex is *not* be a cause of the firing. For if the (counterfactually) male employee were still sexually attractive to his employer, then the employer may very well behave in exactly the same way and thus fire his employee for rejecting his sexual advances all the same. The same argument may apply to the employer's mental representations of the employee as sexually attractive and as sexed female. For so long as they are distinct, so the argument goes, the former should be held fixed in testing for the causal relevance of the latter.

This causal analysis therefore quite plausibly generates the wrong verdicts in this case of sex harassment. This is the first horn of the dilemma. The way to avoid it is to argue that sexual attraction is a part of how sex acts as a cause. But this claim, as I have argued, cannot be *established* by the causal analysis of sex but rather must be *presupposed* by it. But once that claim is in place—once we grant the presumption that actions caused by sexual attraction are actions caused by sex—the causal analysis itself becomes superfluous. This is the circularity of the second horn.[22]

Now consider Appearance Guidelines. Sasha, a female employee, refuses to wear a dress and heels and as a result, is fired for being in violation of her work's appearance guidelines. Can the causal account of sex discrimination in fact account for this result? It might now seem that such firings are not caused by sex as we might have initially supposed. For proper delineation of the various causal factors on scene might point us

[22] What if the employer's representation of the employee as sexually attractive is caused by his representation of her as sexed female? For one, not all cases will have this structure, but even granting it: there still remains the question of whether to control for sexual attraction in determining sex's causal relevance, and I claim there is no question-begging way to answer this question. My discussion of the case Implicit Bias below will explain why.

to a causal contrast in which a *man* in violation of a set of appearance guidelines that applied to him is also fired. So, it seems that it is not Sasha's sex status but her violation of a policy, which could apply to any employee of any sex status, that causes the firing. Certainly, the appearance guideline in this case *makes reference* to sex. But insofar as being sexed female and being in violation of the appearance guidelines are delineated as two distinct causal factors, the latter should be controlled for in the right causal contrast. An analogous argument can be made about the clear distinction between a mental representation of Sasha's sex and a mental representation of their violation of the policy. Just as the harasser's sexual motivations may be fixed across the two scenarios in the case of sexual harassment, why shouldn't the employer's motivating belief that Sasha is in violation of the appearance guidelines also be fixed in this case?

Finally, how about Implicit Bias? Ameera's manager's perception of her as sexed female affected how he evaluated her responses to his questions. By stipulation, his perception of Ameera *as brusque* is brought along by his perception of her as being sexed female. But the mere fact that the former representation depends on the latter, be it causally or non-causally, does not settle what the right causal contrast is for determining whether Ameera's manager discriminates on the basis of sex. For there remains the question of proper causal delineation: whether the representation of Ameera as brusque should be controlled for when assessing the effect of sex on the manager's evaluation.

Suppose first that the perception of Ameera as brusque is a factor that is causally distinct from the perception of her sex status. On this view, it should indeed be controlled for when drawing out the causal effects of sex. In this case, the manager does not discriminate on the basis of sex, since he is put off by what he perceives as brusqueness all the same. This verdict may be avoided only if one stipulates that the causal effects of the manager's perception of Ameera as brusque cannot be cleanly delineated from those of his perception of her sex status. But this is what the causal analysis is itself supposed to *reveal*: granting that the manager's perception of Ameera as brusque caused his poor evaluation of her, does his perception of Ameera as sexed female cause that outcome, too? Once again, one may deny the (presumably wrong) verdict that such cases are not discriminatory on the basis of sex, by claiming that effects of brusqueness *are* effects of sex—though on pain of circularity.

Finally, what about a framing of Implicit Bias where the manager's perception of Ameera as brusque is *causally downstream* from his perceiving her to be sexed female?[23] For one, stipulating a causal link between the two entails that they are distinct, since

[23] For an explicit articulation of this reading of cases of implicit bias, see Shin 2010. A similar account appears in Eidelson 2015, pp. 20–24.

causal relations are typically taken to stand between distinct relata. But whether that is so is one of the questions at stake here, so starting with this presumption is problematic, at least without further defense. Still, even granting this causal structure, the causal analysis of sex still finds itself caught in circularity, though in a slightly more roundabout way (no pun intended). A causal contrast that allows brusqueness to vary alongside sex effectively denies that brusqueness is causally distinct from sex, since on this approach, its causal effects automatically count as sex's causal effects, by transitivity. So it, too, begs the question. The alternative option, which holds brusqueness fixed in order to home in on a "direct" effect of sex on the treatment, generates the "wrong" verdict that the case is not an instance of discrimination on the basis of sex. So the dilemma above again rears its head.

No doubt that much more remains to be said about these three cases and how a causal analysis of them can proceed in light of the challenges presented, and I certainly do not claim to have put the matter to rest. By my aim in this section has not been to show definitively that no account of sex and sex causation can be made to work to escape the dilemma. Rather, I have proceeded in the hopes of simply unsettling the widespread assumption in the literature that a standard causal analysis may easily account for these "classic" cases of discrimination. Probing how an analysis of the causal effects of sex actually proceeds in these cases brings out a crucial stumbling block of these analyses: the causal delineation problem. When cases of (presumptive) sex discrimination are based on factors that *refer* to or are *related* to sex—as many "classical" cases of sex discrimination are—the causal delineation problem is a highly non-trivial problem. Draw the line one way and end up vindicating the wrong verdict; draw the line the other way and end up begging the central causal question at hand.

IV. TOWARDS A NORMATIVE-CAUSAL ACCOUNT OF DISCRIMINATION

I have thus far sought to draw out the substantial hurdles that loom ahead of a causal account of discrimination. To have practical value, the analysis must be capable of deciding among competing causal contrasts on principled, non-circular grounds. The preceding arguments cast doubt on prevailing approaches' prospects of clearing this bar.

One might be tempted at this juncture to take the off-ramp from causation entirely. I am not wholly unsympathetic to this response. But it is worth pausing to appreciate the extent to which dispensing with all causal analysis sends us back to the drawing board on several key parts of a theory of discrimination. For one, without causation, we will need to find another explanation for the trademark "because of" clauses in discrimination talk, which link the action at issue with the relevant statuses. This will also involve a major revision of several of the currently leading theories of discrimination, which rely on causation.[24] Meanwhile, a return to motivating reasons, the other contender account for the link condition, is also beset with substantial hurdles. To the extent that a unified account remains a key desideratum of philosophical work on discrimination, the motivating reasons approach is notably lacking and struggles to explain more subtle non-attitudinal or "structural" cases of discrimination and account for what ties these together with the "simpler" interpersonal cases. And while such cases as Sexual Harassment, Implicit Bias, and Appearance Guidelines were once seen as more marginal to the phenomenon of discrimination, they are now widely recognized to be as paradigmatic of discriminatory harms as are cases of deliberate exclusion. Failing to explain cases so established in the discrimination canon can no longer be taken as simply drawing out a theory's consequences; ruling them out now risks disqualification. The question then is: Can an analysis of discrimination account for these alongside other strands of discrimination without at all appealing to sex, race, etc., causation?

Without ruling out the possibility of a non-causal account of discrimination, and with full recognition that further elaboration and finessing of prevailing causal accounts may yet hold promise, I want to ask in this section: What different approach might there be? I want to propose that we might yet be able to preserve a key role for causation in an analysis of discrimination—*without* claiming to wholly overcome the charge of circularity. Circularity is, after all, not always defeating; to the contrary, it can be, as coherentists have liked to say, "virtuous, not vicious." Might there be a way of elaborating a causal account of discrimination that contains a *non-vicious* circle? My aim in the remainder of this article is to sketch one possibility in this vein—call it a "recipe" if you will—and discuss what promise and pitfalls it might present for our theorizing of discrimination.

The idea is this: In the spirit of reflective equilibrium, we might draw on cases of sex causation exhibited in paradigmatic cases of sex discrimination as "data" to formulate a general theory of sex causation, which may then be used to establish the status of more uncertain cases. On this approach, paradigmatic cases of sex discrimination *qua* paradigmatic cases of sex causation sit alongside the standard stock of intuition-based cases that populate the causation literature, involving preemption, overdetermination, trumping, etc.[25] In fact, it is rather natural to think that a wider set of cases than the

[24] See fn. 2.

[25] Those familiar with the philosophy of causation literature will know well the infamous exploits of Billy and Suzy who variously cast their rocks to shatter bottles and windows. These vignettes provide the standard stock of cases against which theories of causation are tested.

usual vignettes needs to be brought to bear on an account of causation featuring social statuses such as sex. But unlike in the standard approach to theorizing causation, the role of these cases is not primarily to serve as test cases that a plausible analysis of causation must be able to vindicate. Instead, they are inputs into an account of what it is for sex to be a cause, that is, an account of the line that separates those effects that constitute sex's causal effects from those that are distinct from and thus may confound it. As I have argued, how exactly this line is to be drawn is the crux of dispute in controversial cases of sex discrimination. My proposal is that we may draw on our intuitions about paradigmatic cases of sex discrimination qua paradigmatic cases of sex causation—for example, our intuition that cases like Sexual Harassment, Implicit Bias, and Appearance Guidelines in addition to interpersonal animus-driven cases are bona fide cases of sex discrimination and thus cases of sex causation—to answer the causal delineation problem. From here, this account of the line that distinguishes sex's causal effects from those might compete with it can be applied to deliver verdicts on cases of sex causation where our intuitions are less firm—for example, to decide between Sex+ Cases and Sex+ Cases*. Causal theorizing, on this approach, is shot through with normative theorizing insofar as the data from which an account of sex's causal effects is developed contains normative inputs; thus call this a *normative-causal* approach to discrimination.

This approach to analyzing discrimination embraces the methodological view that our ethical judgments of what is discriminatory on the basis of sex and race gives us insight into what sex and race causation *are*. And if these notions of causation are themselves informed by our normative judgments about discrimination, then an account of what is discriminatory on these bases that makes use of such causal notions will certainly fail to be reductive. Still, it may not be viciously circular, if the causal judgments to which the account appeals in an explanation of whether a given case is discriminatory do not derive from the case in question. On this way of proceeding with a causal analysis of discrimination, what is needed is indeed information about discrimination and causation—but it is information about *other* cases to determine the status of this one. This contrasts with the vicious metaphysical circularity that besets extant causal accounts, which attempt to ground an answer to whether a given act constitutes sex discrimination or causation in an account of sex causation that presupposes the causal status of the very act in question.

Thus, the kind of circularity in the normative-causal account can be likened to that which lies in the leading interventionist theory of causation itself. James Woodward's defense that such circularity is non-vicious might here be instructive. Despite the fact that an interventionist defines what it is for X to cause Y in terms of an intervention I, causal routes between *X* and *Y*, and other causes of *Y*—themselves all causal notions—the interventionist theory is "not viciously circular in the sense that the characterization of an intervention on *X* with respect to *Y* itself makes reference to the presence or absence of a causal relationship between *X* and *Y*"; rather, the interventionist theory characterizes the causal relationship "by appealing to facts about *other* causal relationships."[26] All that is needed in order to accept such an account is that, as Woodward puts it, "[w]e begin in *media res*… with a stock of already known causal and correlational information and use this to reach new conclusions about other causal relationships perhaps using these new conclusions to revise other previously accepted causal beliefs."[27] By analogy, all we need in my proposed normative-causal account of sex discrimination is that the account of how sex acts as a cause applied to a given case of disputed sex discrimination is developed from analysis of *other* cases of sex causation, which are less controversial.

This is, of course, only a sketch of how such an account might work; a fuller analysis will have to wait for future writings. But to give a better sense of why such a project might be worth pursuing, I want to turn to say more about the prospects for a normative-causal approach to discrimination and the hurdles that I take accounts developed in this vein to face.

I will start by noting that although the thought that causal facts might be sensitive to normative ones might seem at first downright heretical, it in fact accords well with views that have wide support in the philosophical literatures both on causation and on race and gender. Many prominent theories of causation to varying degrees make explicit room for normative notions. These include accounts which explicitly note that in some circumstances, causal facts are sensitive to moral and political ones.[28] Scholars of race and gender, meanwhile, commonly take an approach to theorizing that freely mixes the metaphysical and the ethical. In particular, substantive normative considerations have been argued to impinge even on our so-called descriptive projects when it comes to elaborating an account of, say, "racism" or "woman."[29] More broadly, to the extent that normative phenomena like sex discrimination are among the key explananda

[26] Woodward 2003, p. 104, 105.

[27] Woodward 2008, pp. 205–206.

[28] For an argument that the causal domain is sensitive to the moral domain, see McGrath 2005. For a focus on social and political concerns related to race and sex, see Hu forthcoming.

[29] See, for example, Tommie Shelby's discussion of the various normative considerations that figure in deciding on a wide-scope versus narrow-scope conception of racism in Shelby 2014. Esa Díaz-León (2016, pp. 245–258) shows how "moral and political considerations can be relevant to the descriptive project of finding out what certain politically significant terms actually mean."

that a good social metaphysics will be able to explain, it is not so far-fetched that our metaphysical theories of sex and of sex causation might themselves be laden with ethical considerations. This explains how a normative-causal account of discrimination may stay in keeping with a standard causal-explanatory interpretation of the hallmark "on the basis of" and "because of" locutions in discrimination talk—while also making sense of why actions "on the basis of" or "because of" sex are often morally wrongful. For, as it turns out, what it is for *sex* to causally explain some outcome is itself in part a normative matter.

A normative-causal account does, however, mark a substantial departure from going theorizing in discrimination theory itself. In particular, admitting a normatively-inflected analysis of sex's causal effects results in a picture of the concept that is at odds with a principal tenet of philosophical theories of discrimination. Insofar as sex causation is an element of the *concept* of sex discrimination, that concept can only be so "non-moralized." This narrows the gap somewhat between a theory of discrimination and a theory of *wrongful* discrimination as pertains to certain social statuses—a gap that most scholars of discrimination implicitly or explicitly endorse.[30] In particular, an analysis that allows first-order ethical judgments about racial or sex-based wrongs and harms to figure directly into one's theory of race and sex causation and discrimination invites further debate about the *kinds* of first-order ethical judgments that are permitted to enter the normative-causal account. This will likely turn on one's moral theory of discrimination. So, a view that takes the wrongness of discrimination to lie in the wrongness of, say, disrespect (Eidelson) may easily recognize a given case of disrespect as discriminatory and on that basis, take it to be a paradigmatic case of sex causation. Whereas a different moral theory which takes the wrongness of discrimination to inhere in, say, demeaningness (Hellman) might be less certain about the same case and so not include it among such core cases from which to draw out a broader theory of sex causation and discrimination. So, different accounts of what is wrong in wrongful discrimination may endorse different cases as "core" to an analysis of sex or race causation and discrimination, depending on the extent to which they exemplify the wrong in question.[31] This means that different moral analyses

[30] That said, the normative-causal approach still allows that a case may be deemed discriminatory where such a determination indeed involves moral analysis, without ruling that the case is all-things-considered morally wrong. For example, affirmative action may mark one such case. I thank a reviewer at *Political Philosophy* for this suggestion.

[31] I have here mentioned only monist theories of wrongful discrimination, but I should also note a different consequence of taking on a pluralist theory of wrongful discrimination. A normative-causal account of discrimination which permits a variety of ethical judgments to figure

will have wider consequences than has been typically presumed, as they lead also to different accounts of race or sex causation and discrimination more broadly. These connections between what discrimination on the basis of race or sex *is*, what makes such discrimination *wrong*, and what it is for race or sex to be *causes* challenges the predominant philosophical methodological approach to these topics, which by and large treats them separately.

Of course, this very entanglement comes with its own set of theoretical and practical risks.[32] As a general matter, reflective equilibrium as a method for moral theorizing always runs the risk of overfitting to pre-existing intuitions. But it faces a distinctive methodological pitfall when used to develop a philosophical account of discrimination on the basis of sex (or race, etc.). For our intuitions about which cases are genuine instances of discrimination on the basis of sex may draw upon our views as to whether they ought to be *legally* deemed and thus *legally* prohibited as discriminatory on the basis of sex. But to the extent that these latter intuitions are influenced by what are presumably theoretically extraneous matters—matters such as how certain jurisdictions have come to structure their discrimination doctrine—they are suspect as intuitions about sex discrimination *as such*.[33] On the other side of things, when used to develop a *legal* analysis of what constitutes discrimination on the basis of sex, the normative-causal account's open appeal to substantive judgments about the moral

into one's theory of race and sex causation risks blurring the line separating discrimination from a larger pool of wrongs and harms related to race and sex. For example, suppose that a judgment that an action reinforces a pillar of patriarchy or the subordination of women is permitted to ground a judgment of sex discrimination and so sex causation. A normative-causal account of sex discrimination which starts here might end up being so expansive as to lead one to wonder what makes actions distinctively *discriminatory* on the basis of sex as opposed to being just generically bad for women? If marking out an action as discriminatory amounts to little more than pointing out that it belongs to a large family of ethical shortfalls relating to a given social category, the notion itself appears redundant. That said, this may well not be considered a defect of a theory of discrimination. Notably, Moreau's (2020a) pluralism about wrongful discrimination is deeply tied in with the task of making sense of the "expanded conception of discrimination," which encompasses many group-based harms and wrongs that by her own lights are new additions to our understanding of discrimination. It thus stands to reason that Moreau herself would at least be less ill at ease with such expansionary consequences.

[32] The following remarks are greatly indebted to a reviewer at *Political Philosophy.*

[33] For instance, it might be that the particular list of protected characteristics covered by anti-discrimination law within a given legal system exerts pressure on our general intuitions about what constitutes discrimination on the basis of what characteristics.

wrongness of cases runs up against legal ideals of neutrality and objectivity. This in turn raises questions of what exactly it is to uphold such ideals in judicial reasoning about discrimination and whether doing so is even possible let alone desirable.[34] Suffice to say, all this leads to an entirely different realm of highly complex and knotty issues.[35]

Drawing out these risks and revisionist implications is a valuable exercise. For if, as I suspect, they are what drive theorists of discrimination away from a normative-causal account, they indicate theoretical pressure points in analyses of discrimination that have shaped how the literature has developed. In so doing, they clarify the stakes of analyses of discrimination, allowing us to better appreciate the lines of inquiry that these commitments enable as well as those they close off, and so in turn allowing us to better scrutinize whether we *should* continue to cling onto these features or give them up in exchange for a theory with other theoretical and practical benefits.

I want to now close this paper with some remarks on why I think the aforementioned theoretical costs *are* worth paying. They are, I claim, because what we get in exchange is an analysis of discrimination that better than prevailing accounts, successfully integrates the core social and ethical dimensions of the phenomenon.

V. CONCLUSION

A normative-causal approach to discrimination vindicates the longstanding thought that what connects a given act of discrimination to the grounds of discrimination is a causal link—and crucially, it does so in a way that *explains why* the cases raised in Sex+ Cases are genuinely troublesome (and not just due to "mistakes" in the discourse). Furthermore, it suggests that what we should say about charges of question-begging in the literature is that they are both apt, because true, and also inapt, insofar as they suggest falsely that there are wholly non-circular ways to a solution. I take these to be substantial theoretical virtues of an account of discrimination.

Still, in my view, the most important advantage of a normative-causal approach over extant analyses lies elsewhere: in its integration of the social and ethical dimensions of discrimination. The fact that the features of conventional discriminatory concern such as sex and race are markers of *social difference* is typically taken to be significant only for normative analyses of discrimination. But it seems to me that the social nature of these

[34] For a classic discussion of courts' attraction to interpretations of antidiscrimination law that abide by ideals of legal neutrality and objectivity, see Fiss 1976.

[35] A realm discussed in splendid detail in Minow 1987.

cases of discrimination has significance not just for our understanding of what makes them wrong but for our *conceptual* understanding of these kinds of discrimination. As markers of social difference, race and sex track not just differences in skin color, phenotype, chromosomal makeup, secondary sex characteristics, and so on, but also differences in the distribution of social goods: resources, norms, expectations. These differences are systematically bundled together such that the causal effects of race and sex are now broadly taken to include effects issuing from the latter set of factors, which lie outside of what might traditionally be taken to be the *sine qua non* of the category. To the extent that the boundaries of these categories and their causal effects are under contestation due to substantive ethical debate—Should sex include also gender conforming behavior? Which ones? Are these effects thereby effects of sex?—it seems only natural that what it takes to discriminate on the basis of these categories would also be a normative matter.

A normative-causal approach presents a promising avenue to further elaborate this thought. It is by now a truism that discrimination is a "social problem," that it does not take place against a "background" of social inequality but actively produces and reproduces that inequality, and that we fail to understand it fully so long as we apply to it the framework of one-off animus-fueled interpersonal transactions. These features are, in my view, worth placing at the core of the discrimination concept. They guide us towards an analysis of what discrimination *is* that attends closely to how various practices reflect, make, and remake our unjust social reality—a story which will most certainly be both normative, to pick out that unjust social reality, and causal, to pick out those processes that make it.

ACKNOWLEDGEMENTS

This paper owes its existence to years of conversation on this and related topics in discrimination theory with Issa Kohler-Hausmann. It was subsequently rethought and honed through discussions with participants at the Princeton MAP talk series and the Philosophy Mountain Workshop. I am also grateful to the thorough and thoughtful comments I received from two anonymous reviewers at *Political Philosophy*, which undoubtedly greatly improved the final product that appears here.

COMPETING INTERESTS

The author declares that she has no competing interest.

REFERENCES

Àsta. 2018. *Categories We Live By: The Construction of Sex, Gender, Race, and Other Social Categories.* Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780190256791.001.0001.

Berman, Mitchell N., and Guha Krishnamurthi. 2021. Bostock was bogus: textualism, pluralism, and Title VII. *Notre Dame Law Review*, 97: 67.

Dembroff, Robin, and Issa Kohler-Hausmann. 2022. Supreme confusion about causality at the supreme court. *CUNY Law Review*, 25: 57.

Dembroff, Robin; Issa Kohler-Hausmann; and Elise Sugarman. 2020. What Taylor Swift and Beyoncé teach us about sex and causes. *University of Pennsylvania Law Review Online*, 169: 1.

Díaz-Léon, Esa. 2016. Woman as a politically significant term: a solution to the puzzle. *Hypatia*, 31 (2): 245–258. https://doi.org/10.1111/hypa.12234.

Eidelson, Benjamin. 2015. *Discrimination and Disrespect*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198732877.001.0001.

Eidelson, Benjamin. 2022. Dimensional disparate treatment. *Southern California Law Review*, 9: 785.

Fiss, Owen M. 1976. Groups and the Equal Protection Clause. *Philosophy & Public Affairs*, 5 (2): 107–177.

Greiner, James, and Donald Rubin. 2011. Causal effects of perceived immutable characteristics. *Review of Economics and Statistics*, 93 (3): 775–785. https://doi.org/10.1162/REST_a_00110.

Haslanger, Sally. 2019. Tracing the sociopolitical reality of race. Pp. 4–37 in *What Is Race?: Four Philosophical Views*, ed. Joshua Glasgow, Sally Haslanger, Chike Jeffers, and Quayshawn Spencer. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780190610173.001.0001.

Hellman, Deborah. 2008. *When is Discrimination Wrong?* Cambridge, MA: Harvard University Press. https://doi.org/10.4159/9780674033931.

Hellman, Deborah. 2023. Defining disparate treatment: a research agenda for our times. *Indiana Law Review*, 99 (1): 206–207.

Hu, Lily. Forthcoming. Normative facts and causal structure. *Journal of Philosophy*.

Lippert-Rasmussen, Kasper. 2013. *Born Free and Equal?: A Philosophical Inquiry into the Nature of Discrimination*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199796113.001.0001.

McGrath, Sarah. 2005. Causation by omission: a dilemma. *Philosophical Studies*, 123 (1–2): 125–148. https://doi.org/10.1007/s11098-004-5216-z.

Minow, Martha. 1987. Foreword: justice engendered. *Harvard Law Review*, 101: 10.

Moreau, Sophia. 2020a. Equality and discrimination. Pp. 171–90 in *The Cambridge Companion to the Philosophy of Law*, ed. John Tasioulas. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781316104439.

Moreau, Sophia. 2020b. *Faces of Inequality: A Theory of Wrongful Discrimination*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780190927301.001.0001.

Shelby, Tommie. 2014. Racism, moralism, and social criticism. *Du Bois Review: Social Science Research on Race*, 11(1): 57–74. https://doi.org/10.1017/S1742058X14000010.

Shin, Patrick S. 2010. Liability for unconscious discrimination: a thought experiment in the theory of employment discrimination law. *Hastings Law Journal*, 62: 67.

Singh, Keshav, and Daniel Wodak. 2024. Does race best explain racial discrimination? *Philosophers' Imprint*, 23 (24). https://doi.org/10.3998/phimp.2463.

Weinberger, Naftali. 2022. Signal manipulation and the causal analysis of racial discrimination. *Ergo*, 46: 1264–1287. https://doi.org/10.3998/ergo.2915.

Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press. https://doi.org/10.1093/0195155270.001.0001.

Woodward, James. 2008. Invariance, modularity, and all that: Cartwright on causation. Pp. 210–249 in *Nancy Cartwright's Philosophy of Science*, ed. Luc Bovens, Carl Hoefer, and Stephan Hartmann. New York: Routledge. https://doi.org/10.4324/9780203895467.

Zatz, Noah D. 2017. Disparate impact and the unity of equality law. *BU Law Rev*, 97: 1357.