Consensual Discrimination

Andreas Bengtson and Lauritz Aastrup Munch

Abstract: What makes discrimination morally bad? In this paper, we discuss the putative badness of a case of *consensual* discrimination to show that prominent accounts of the badness of discrimination—appealing, *inter alia*, to harm, disrespect and inequality—fail to provide a satisfactory answer to this question. In view of this, we present a more promising account.

Forthcoming: *The Philosophical Quarterly*

1. Introduction

What, if anything, makes discrimination morally bad? In the literature, we find several answers such as *harm* (Lippert-Rasmussen 2013), *disrespect* (Eidelson 2015), or *inequality* (Moreau 2020).

In this paper, we show that prominent accounts cannot predict and explain at least one type of badness that we find in some cases of discrimination. This suggests that these accounts provide us with an incomplete picture of the normative profile of discrimination. In response to this, we sketch an alternative account focusing on the badness of acting on motivating reasons that lack reason-giving force.

Our argument begins with the observation that the badness of discrimination appears to persist even if it is consented to by its victim. This observation is partly motivated intuitively (consensual discrimination *seems* morally problematic). It is further motivated by the observation that few jurisdictions permit legal claims against discrimination to be waived via consent. But what if you deny what we take for granted in this paper, that consensual discrimination is normatively deficient? (And so deny that an adequacy constraint on a successful account of the badness of discrimination is that it predicts that consensual discrimination is problematic.) You should read on even still to learn what challenges such a view would face.

2. Consensual and non-consensual discrimination

Consider the following pair of cases:

> *Firing*. Boss fires Employee because she is a female. Had Employee been a male, she wouldn't have been fired.

> *Firing #2*. Identical to *Firing* but Employee (validly) consents to being fired because of her gender.[1] Had Employee been a man, or had Employee not consented, Boss wouldn't have fired her.

Firing is a paradigmatic example of wrongful discrimination according to most contemporary accounts (Lippert-Rasmussen 2013; Eidelson 2015; Moreau 2020): Boss treats Employee disadvantageously and this is best explained by Employee's (or Boss's perception of Employee's) gender. We should regard this verdict as a datum. Any minimally plausible account of discrimination should generate the verdict that *Firing* is a case of wrongful discrimination.

How about *Firing #2*? In this case, Employee issues her consent to being treated disadvantageously based on her gender, and Boss's conduct is appropriately sensitive to whether Employee consents. This means that Boss's conduct in *Firing #2* is partly explained by Employee's gender and partly by her consent. We can gloss this using counterfactual dependence relations: had it either been the case that Employee didn't consent, or had it been the case that Employee were a man, Boss wouldn't have fired her.[2] This is different from what goes on in *Firing*, where Boss wouldn't be responsive to Employee's consent, and where Employee doesn't in fact consent.[3]

[1] Some may believe that this is simply a result of *adaptive preferences* (for discussion of adaptive preferences, see Enoch 2020). We assume that *Firing #2* involves genuine consent. This is clearly a conceptual possibility, and that suffices for our purposes.

[2] Notice that there is no asymmetrical counterfactual dependence in terms of Boss's responsiveness to consent: Had Boss taken themselves to have reasons to fire a manly version of Employee (reasons independent of gender), they would also only have done so had Employee issued their consent. That is to say, there is no charge of differential treatment in terms of responsiveness to consent.

[3] What about a case where Employee in fact consents but Boss isn't responsive to this consent? We'll set this case aside here, because it looks morally defective to act in the presence of consent without being guided by it. To test for the significance of consent to discrimination, we need to consider a case where we can be confident that the discriminator is suitably guided by the discriminatee's consent.

So that's how *Firing #2* differs from *Firing*. How should we evaluate it, normatively speaking? Perhaps *Firing #2* is less objectionable than *Firing* in some respects. But it seems hard to accept that everything is just fine. Even if Employee is not wronged by Boss, it seems clear enough that there is still something normatively deficient about Boss's conduct. We can appreciate this by noting how Boss's conduct—despite being responsive to the consent of Employee—is still sensitive to Employee's gender. Had Employee been a man, Boss wouldn't have considered firing him.

If true, at least one normative flaw of discrimination doesn't go away even if consented to by the discriminatee.

We shall take this as a datum, but it's possible to say something in defense of this point apart from intuitions. For instance, many anti-discrimination laws do not include consent as a basis for making otherwise legally problematic forms of discrimination legally permissible. This would seem to suggest that if you judge that consent purges discrimination of its characteristically problematic feature, then you incur the cost of explaining how, suitably interpreted, these bodies of law can be said to regulate discrimination in a way that is faithful to its normative profile.

It's worth pausing to clarify what we mean by consent, and what we mean by saying that it doesn't remove the badness of discrimination. The characteristic function of consent is to give people control over the duty of others (Bolinger 2019; Dougherty 2021; Hurd 1996). For example, Lou is not permitted to touch Morag, but if Morag consents to this (e.g., by uttering the words "please, touch my body"), Lou can now permissibly touch Morag's body. In this way, consent creates permissions where there would otherwise exist prohibitions as a default. Moreover, in this specific case, consent purges the act of touching Morag's body of whatever feature that would otherwise make it morally problematic. It's important to notice that consent can affect the moral status of an act without rendering it non-objectionable. For an illustration, it seems morally wrong to pointlessly torture a pet turtle even though its owner consents to it. In this case, consent may ensure that the owner is not wronged, but it doesn't make the act of torturing the turtle non-objectionable, and it hardly renders it permissible. So, when we say that consent doesn't seem to make discrimination in *Firing #2* non-objectionable, it might be that Employee's consent fails to grant Boss a permission to fire Employee because of her gender. Or it might be that Employee's consent grants Boss permission to fire Employee, but Boss's act might still be stained, normatively speaking. As Hurd (1996: 123) puts it,

> consent can generate a permission that allows another to do a wrong act … it [consent] does not morally transform a wrong act into a right act, but it grants another a right to do wrong. It conveys, in these circumstances, a "stained permission," for the act done remains, in some sense, wrong, and hence, morally stained, but the consent defeats any rights on the part of others (including the person consenting) that the actor not do the wrong act.

Our suggestion is that we can learn something important by tending to the question of whether a type of wrong or normative deficiency is fully removed by consent. If the answer is negative—as we suggest it is in *Firing #2*—it follows that the normative status of the act is not fully under the direct normative control of its victim.[4] This allows us to infer that at least one type of deficiency that is characteristic of discrimination cannot be one that is under the direct control of the discriminatee. It turns out that this insight is a powerful discriminating (no pun intended) tool for making progress on the question of what may be characteristically normatively deficient about discrimination. We'll see that in the next section, where we'll show that four prominent accounts of what's bad about discrimination cannot explain what goes wrong in cases of consensual discrimination.

3. Four accounts of the badness of discrimination

[4] Notice that there can be *directed* wrongs that aren't subject to the normative control of their victims (this is important if you think discrimination is a directed wrong). I owe it to *you specifically* not to enslave you, no matter what you say (Jonker 2020).

3a. The harm account

According to an influential account of the badness of discrimination, the distinctive moral flaw of discrimination is that it's harmful (Lippert-Rasmussen 2013). But that can't at least be the whole story if a flaw of discrimination persists even when consensual. This is because any plausible account of harm should allow that harming can be permissible (without any moral residue) if it's consensual. As Tadros (2011: 23) puts it,

> One thing that can render harm to another permissible is consent. But there are also circumstances in which it has been regarded wrong for one person to harm another even though the other consented to be harmed.

Tadros says that consent sometimes, but not always, can make harm to another permissible. Perhaps the exceptions Tadros has in mind can aid proponents of the harm account in explaining what goes wrong in cases of consensual discrimination? Tadros says there are exceptions if the person harmed has a self-regarding duty to protect themselves from harm. If so, they cannot render harm done to them permissible through their consent. But the exceptions Tadros has in mind won't aid proponents of the harm account. The cases he has in mind are cases of extreme harm, such as outright killings, or acts that severely undermine bodily functions necessary for flourishing (see also Shaber, 2020: 87-88). The profile of these cases doesn't fit with the commonsensical verdict that minimally and moderately harmful forms of consensual discrimination are problematic. This means that the harm account cannot explain why *Firing #2* is normatively deficient.[5]

[5] Perhaps some would want to go further with the self-regarding duty stuff mentioned above and say that the flaw in *Firing #2* is that, by consenting to being discriminated against, Employee violates a duty she has to herself, that is, she wrongs herself. However, this explanation of what goes wrong in *Firing #2* is limited. After all, most philosophers don't believe we have duties to ourselves. We might have hoped for something less controversial to explain the flaw in *Firing #2*. Perhaps this explanation would suffice if we didn't have anything else. But, as we will explain later, we don't need to appeal to duties to self to explain why there is a flaw in *Firing #2*.

An unqualified harm-based account won't work. But for all we've said, a more sophisticated one might. Lippert-Rasmussen's (2013: 165) harm-based account, for instance, is "desert-prioritarian." There are two further components to this account. First, there is the familiar prioritarian idea that the marginal moral value (or disvalue) of gaining some unit of well-being (or ill-being) depends on how well off somebody already is. Second, Lippert-Rasmussen says that desert is a constituent of the base that determines whether a person is worthy of greater or lesser priority. On the resulting view, discrimination is bad when and because it is harmful, and the harm is not allocated in accordance with a desert-prioritarian principle. Can this view pick out what is bad about consensual discrimination?

Notice first that there's no necessary relation between whether a harmful act is consensual and whether the harm is allocated in accordance with desert-prioritarian principles. To illustrate how these things come apart in one direction, imagine that discriminating against a single person based on their race is the option that uniquely satisfies desert-prioritarian principles (suppose this person is extremely well off and extremely deserving, but if we don't discriminate against this person, many other people who are less well-off and much more undeserving will be harmed). We could imagine that this person consented to discrimination, or we could imagine that they didn't. This is significant because it suggests that there can be cases of consensual discrimination that could remain objectionable because the harms that are part of the discriminatory treatment are not distributed in a way that is consistent with desert-prioritarianism.[6] So it seems that on the face of things, Lippert-Rasmussen's account can pick out something that goes wrong in at least some cases of consensual discrimination. Despite this, we don't think Lippert-Rasmussen's account is well positioned to deal with the case of consensual discrimination. But there are many moving parts in his desert-prioritarian account, so it takes a paragraph to see this.

[6] This isn't unique to desert-prioritarianism but seems to be the case for any distributive pattern that does not track consent.

First, insofar as Lippert-Rasmussen's account is meant to be 'harm-based' in the specific sense that the badness of discrimination is ultimately explained by its harmfulness, adding a desert-prioritarian component won't obviously undermine the argument we offered above. To see this, notice that a harm that is not allocated in accordance with a desert-sensitive prioritarian principle could still have its badness removed via consent. Or, if a harm arising from discriminatory treatment was consented to, this fact alone could affect the desert base such that the harm would be acceptably imposed. In any case, were we to learn that the particular type of harm that is picked out by Lippert-Rasmussen's account isn't sensitive to consent in the way that harm paradigmatically is, we might come to believe that the account is not harm-based after all.

Lippert-Rasmussen could respond here that this is precisely the point. He could say either that the badness of discrimination lies with the violation of a distributive principle (such as prioritarianism), or he could say that, suitably interpreted, discrimination is normatively deficient because it's underserved (focusing on the desert-part of his view). This view would be saying that while harm is the property that has the potential to be problematic in cases of discrimination, harm is only objectionable if it violates a principle such as prioritarianism or is undeserved. (Notice that this way of thinking about the moral significance of harm is not esoteric; we normally think that harm is problematic only insofar as some further conditions are satisfied, such as the harm being imposed non-consensually).

This gives us two further views (or families of views, really) that merit consideration. Regrettably, though, a full assessment of these possibilities lies well beyond the scope of this paper. So let's confine ourselves to some general remarks. First, the relationship between desert and consent is in the best case complex, and in the worst case deeply controversial. So our focus on consensual discrimination is unlikely to provide a clean argument in either direction.[7] For that reason,

[7] One could say that insofar a person's responsible choices determine their desert (Lippert-Rasmussen (2013) seems to suggest as much), and consent is a responsible choice, consent is a part of the desert base. But since many other things will probably be as well, it's not obvious that desert could pick out what is problematic in cases of consensual discrimination. Consent doesn't seem to align perfectly with one's deserts, as it were.

we'll set this issue to one side given our focus on teasing out how consensual discrimination—without too many further thorny commitments—can help us understand what makes discrimination normatively defective. But let's note in passing that some are very skeptical of the viability of a desert-based explanation of the badness of discrimination, including Lippert-Rasmussen himself (2023).

How about saying that discrimination is bad because it violates prioritarian principles? As we noted above, the explanation that discrimination is problematic because it doesn't align with a prioritarian distribution of well-being has the right form to say what could be bad about some cases of consensual discrimination because consent and priority come apart. In fact, we can generalize this point and say that any distributive pattern that is not consent-sensitive will have the right form to deal with the case of consensual discrimination (e.g., sufficientarian, egalitarian, etc.).[8]

Our concern with this explanatory strategy, however, is a version of a much more general concern that is familiar from the debate about distributive justice, namely that there's something worrisome about endorsing a distributive principle that is insensitive to people's choices (e.g., in the form of consent, as a species of choosing). To put the point as a question: why should we be concerned with a distribution that lacks a certain pattern (be that egalitarian, prioritarian, or sufficientarian)—and by extension be concerned with discriminatory acts that upset our favored pattern—if these acts are consented to by the people? In order to not make this point come off as excessively libertarian in spirit, it's worth noting that a commitment to choice-sensitivity is widespread. For instance, luck egalitarians say that the pattern we have most reason to prefer is choice-sensitive (Cohen 1989; Lippert-Rasmussen 2015). We think reasons such as the above-mentioned should advise against opting for a non-choice sensitive distributive pattern as our favored explanation of what makes discrimination problematic. And a choice-sensitive account, in turn, should allow for consent to upset patterns (and so cannot explain the badness in *Firing #2*).

[8] By contrast, Nozick's (1974) entitlement theory of justice would be an example of an account that is typically seen as addressing the same question as for instance prioritarianism and sufficientarianism but tracks consent (or something very close to it).

In sum, we don't think that Lippert-Rasmussen's desert-prioritarian harm-based account is in a good position to deal with the case of consensual discrimination. That's clearest if we interpret the view as harm-based. But, as we have pointed out, a non-harm-based interpretation may also be problematic, though for reasons that are more complex because they touch upon deeper issues about how best to understand notions such as desert and distributive justice.

3b. The respect account

Another prominent view is due to Eidelson. It says that "discrimination is intrinsically objectionable when it is basically disrespectful of the personhood of those who are discriminated against" (Eidelson 2015: 95; see Thomsen 2023 for extended discussion). To Eidelson, there are "two elemental facets of moral personhood—a person's moral worth and her autonomy—that are proper objects of recognition respect" (96). This corresponds, Eidelson says, to two separable ways in which it's possible to disrespect people's moral personhood.

Let's first consider the part of the view that says that discrimination is disrespectful because it consists of a failure to take people's *moral worth* appropriately into account in one's deliberation. Eidelson substantiates this idea as follows:

> "The key claim in staking out this position is what I will call the *interest thesis*:
>
> > To respect a person's equal value [moral worth] relative to other persons one must value her interests equally with those of other persons, absent good reason for discounting them" (Eidelson 2015: 97).

To make this more concrete, consider an example offered by Eidelson:

> Suppose that Adam is choosing whether to promote Fatima, who is of Arab descent, or Christopher, who is white. Adam considers the benefit to Christopher of getting the promotion, and gives this some weight in his decision. But because Fatima is of Arab descent,

> Adam gives the equal benefit to Fatima of getting the promotion less weight in his decision, no weight at all, or even negative weight. (...) What is disrespectful of Fatima is not the bare fact of discounting her interests, then, but doing so for no good reason (Eidelson 2015: 96-98).

Focus on the last part, the idea that the flaw in discrimination amounts to "discounting interests for no good reason." If we consider cases like *Firing #2*, there at least seems to be a good reason for discounting Employee's interests, namely the reason generated by Employee's consent. If this seems implausible, suppose instead that Employee not only consented to, but also *requested* that, her interests be given less weight. It's hard to see that this request couldn't constitute a reason for giving Employee's interests less weight, and so we could imagine a version of *Firing #2* where Boss would act only if Employee also requested the treatment at stake. It seems to us that this version of the case would remain problematic despite Employee's request.

It might be objected that despite Employee requesting the treatment, Boss nevertheless takes Employee's gender as *a* reason for giving her interests lower weight and this is what's problematic. We agree with this observation. But we're less sure that the right explanation of what goes wrong involves appealing to the wrongfulness of discounting people's interests. One reason why is that people often enjoy direct normative control—for example, via the power of consent—over whether it's appropriate to give their interests a certain kind of weight. We see this well illustrated in cases of paternalism where the characteristic problem is taking the promotion of people's interests as a reason for acting without taking into consideration people's say-so over how their interests are used to justify the action (e.g., Grill 2015; Groll 2012). Paternalism is insulting because it involves the promotion of a person's interests without their consent. Thus, consent, naturally, removes this complaint. If this says something general about the moral relationship between taking people's interests into account when deliberating and the will of these people, then it's hard to believe that we can explain what's wrong about discrimination by appealing to how it involves discounting interests in the face of consent (as in *Firing #2*).

To make a similar point with a different analogy, consider a familiar type of rescue case where two people with equal claims are in peril and only one can be saved. Since each have an equally strong claim to being rescued, a rescuer ought to give them an equal chance (perhaps by flipping a coin). But notice that if one of those in peril decides to waive their equal claim to being rescued (for example, by communicating a refusal to be rescued or by permitting that the other be saved), then this is something the rescuer ought to take into consideration. In such cases, consent (or the exercise of analogous normative powers such as the power to refuse) constitutively affects what weight it's appropriate to give to people's interests.

Taken together, it's not obvious that the first part of Eidelson's account can explain what goes wrong in cases of consensual discrimination. On one interpretation, we might agree that these cases involve discounting people's interests for no good reason but deny that this is what makes the cases problematic in the presence of consent (by analogy to paternalism). On another interpretation, it's not even obvious that consensual discrimination necessarily involves discounting anybody's interests if we think that powers like consent have the capacity to constitutively affect what weight it's appropriate to give to people's interests (by analogy to the rescue case). Accordingly, it seems that the first part of Eidelson's account of what it means to respect personhood cannot obviously capture what goes wrong in *Firing #2*.

The next aspect of Eidelson's theory is the idea that discrimination can be disrespectful because it fails to take into consideration the autonomy of those subjected to discrimination:

> Some discrimination is disrespectful of the victim's standing as a person (...) because it fails to treat her as an individual autonomous being. This is an essential and distinct ingredient in a satisfying account of the morality of discrimination (Eidelson 2015: 128).

But it's easy to see that this aspect of Eidelson's account cannot explain why consensual discrimination is problematic. If you're acting with the *autonomous* consent of the person, it's hard to see how you could fail in treating the person as an individual autonomous being. Morally transformative consent is standardly taken as a direct expression of a person's autonomous will, after all (Dougherty 2021).

In sum, although we're sympathetic to the idea that the distinctive flaw of discrimination consists in a "deliberative failure" of some form—a point to which we return—Eidelson's focus on moral worth and autonomy, as two facets of respecting moral personhood, seems like the wrong starting point because such things (on Eidelson's interpretation of them) seem to be under the control of people via their consent or other autonomy-expressing moral powers.

3c. The inequality account

A third account of the wrongness of discrimination has been proposed by Moreau (2020). According to her, discrimination may be wrong because it involves (i) unfair subordination; (ii) deliberative unfreedom; and/or (iii) denying people access to basic goods. What unifies these, Moreau argues, is inequality. In short, discrimination is wrong when and because it constitutes (wrongful) inequality. Let's explore what the three parts of Moreau's view might say about consensual discrimination.

Starting with the latter—that discrimination is wrong if it denies people access to basic goods—a good is a basic good for a particular person if and only if:

(i) Access to this good is necessary in order for this person to be a full and equal participant in her society; and

(ii) Access to this good is necessary in order for this person to be seen by others and by herself as a full and equal participant in her society (Moreau, 2020: 126)

As an example, she mentions the case of indigenous communities denied access to safe drinking water. It's obvious why access to safe drinking water (in case other groups in society have such access) is necessary to be, and be seen by others as, a full and equal participant in society. But such concerns need not arise in cases of consensual discrimination. Indeed, it's not even clear what the good at stake is in *Firing #2*. In any case, we may simply suppose that *Firing #2* takes place in an otherwise egalitarian society. In that case, her consenting to being discriminated against on behalf of her gender need not threaten her status as a full and equal participant in society. Indeed, denying her this option may threaten her status as a full and equal participant. Note, finally, that Moreau points out that the person must be denied *access to* the good for it to constitute wrongful discrimination. But she may have had access to the relevant good even if she consents to discrimination. So, appealing to lack of access to basic goods cannot explain why *Firing #2* is objectionable qua discriminatory.

Let's turn to the second part of Moreau's account: that discrimination is wrong if it involves deliberative unfreedom. According to Moreau, deliberative freedom is

> the freedom to deliberate about one's life, and to decide what to do in light of those deliberations, without having to treat certain personal traits (or other people's assumptions about them) as costs, and without having to live one's life with these traits always before one's eyes (Moreau, 2020: 84).

Take Black people in a racist society. They lack deliberative freedom. When they deliberate about what to do, they'll have to take their race into account both in the sense that it makes it more difficult for them to pursue certain options—e.g., to get a certain job—but also in the sense that their race has been "made an issue" (Moreau, 2020: 85).

Now, it's clear why Black people lack deliberative freedom in racist societies. And why women lack deliberative freedom in sexist societies. Their race and gender are made an issue when they deliberate about what to do. But, for our purposes, we may imagine that *Firing #2* happens in an otherwise egalitarian society. In that case, the woman consenting to being fired because of her gender does not lack deliberative freedom. She voluntarily chooses to make her gender a part of her deliberation. But there still seems to be discriminatory taint in *Firing #2*. Thus, appealing to deliberative unfreedoms cannot explain why *Firing #2* is objectionable qua discriminatory: it's underinclusive.

Let's, finally, turn to the third part of Moreau's theory: that discrimination is wrong if it involves unfair subordination. According to Moreau, one group is unfairly subordinated to another when:

> (i) The members of that group have, across a number of social contexts, less relative social and political power and less relative de facto authority than the other group; and
>
> (ii) The members of that group have, or are ascribed, traits that attract less consideration or greater censure across a number of different social contexts than the corresponding traits of the empowered group; and
>
> (iii) These traits are the subject of stereotypes, which help to rationalize the differences in power and de facto authority, the habits of consideration and censure, and the structural accommodations; and
>
> (iv) There are structural accommodations in place in society that tacitly accommodate the needs of a superior group while overlooking the needs of at least some members of the subordinate group; and these accommodations work together with stereotypes to rationalize the differences in power and de facto authority and the differences in consideration or censure (Moreau, 2020: 62)

In the context of explaining why *Firing #2* is objectionable qua discriminatory, Moreau's account of unfair subordination suffers from the following dilemma. Either discrimination is objectionable because it constitutes unfair subordination, or discrimination is objectionable when it takes place in a context in which there is unfair subordination. Start with the first horn. The problem is that Moreau's account of unfair subordination is posed at almost a structural level where one group is unfairly subordinated to another if it has less relative power and de facto authority and there are structural accommodations in place. But *one* instance of discrimination—as in *Firing #2*—cannot change whether one group is unfairly subordinated to another. If the group of women wasn't already subordinated to the group of men, *Firing #2* wouldn't make it so. It simply might not make any difference to relative power and de facto authority and structural accommodations. So the view that discrimination is objectionable because it's constitutive of unfair subordination cannot explain why *Firing #2* is objectionable qua discriminatory. This is the first horn of the dilemma. Alternatively, Moreau might say that discrimination is objectionable when it takes place in a context in which there's unfair subordination, i.e., when the discriminator is a superior group and the discriminatee is an inferior group. But this will not cut much ice in our dialectical situation. Indeed, as mentioned above, we may imagine that *Firing #2* takes place in an otherwise egalitarian society in which no group is unfairly subordinated to another. This is the second horn of the dilemma.

In sum, Moreau's account of what makes discrimination wrongful cannot explain why *Firing #2* is objectionable qua discriminatory, irrespective of whether we appeal to lack of access to basic goods, deliberative unfreedom or unfair subordination.

3d. The vicarious wrongdoing account

Jonker (2019) has recently argued that (part of) the distinctive flaw in discrimination is found in its *vicarious* nature. Jonker focuses on cases such as the following:

> *Mixed Marriage.* C, a white man married to a black woman, is fired from his job as a basketball coach because his white employer objects to his marriage (Jonker 2019: 209).

As Jonker points out, part of the problem here boils down to how the white man is fired. But another part—the vicarious part—targets the black woman. Focus on the latter idea. Jonker says that this points to a distinctive aspect of the wrongness of discrimination:

> Non-associational cases of discrimination typically involve vicarious wrongs. The disregard of a discriminator is aimed not at the one who is discriminated against, but at the class of people to which she belongs (or in associational cases, is associated with); and the victim of discrimination bears harm that the discriminator wishes to inflict (or neglects not to inflict) upon the group (Jonker 2019: 212).

Admittedly, Jonker's account stops short of being a full account of the badness of discrimination, as it's more concerned with identifying its directionality. But it's worth considering here because a plausible competing explanation of why the primary victim cannot remove a flaw of discrimination through their consent is that this flaw doesn't target the consenting agent, but third parties. On a natural interpretation of Jonker's view, we can locate the vicarious victim of discrimination via the mental states of the discriminator. This enables us to test the relevance of Jonker's argument here. To have a stab at that, consider:

> *Firing #3.* Boss fires Employee not because of her gender, but because of the ethnicity of Employee's partner. Employee and Employee's partner consent to this treatment.

Firing #3 is meant to have a vicarious structure analogous to Jonker's *Mixed Marriage* but is otherwise intended to remain as similar as possible to *Firing #2*.[9] For our purposes, we only need to consider whether the consent of the partner changes much in terms of how we should evaluate the case. And on this, we're inclined to think that *Firing #3* looks as deficient as *Firing #2*. Of course, a possible explanation of this could be that the posed vicarious victim of discrimination, suitably interpreted, isn't a genuine victim which in turn could explain why their consent lacks relevance. But since we're sympathetic towards this idea in general, and since we must take the possibility of vicarious victims of discrimination for granted to test the relevance of this idea here, we think the best interpretation is in line with what we've assumed so far: There's a sense in which consent is irrelevant for determining the normative status of discriminatory acts.

4. Flaw in responsiveness to reasons

We take it to be a significant result that leading accounts of what makes discrimination objectionable have a hard time making sense of consensual discrimination. In the rest of the paper, we'll sketch an alternative that fares better on this point. First, we'll consider what it is about *Firing #2* that seems unaffected by consent. Second, we'll develop an account capable of explicating that feature of *Firing #2* as a normative flaw. Third, we'll address directly why this account seems well-motivated by the result that consensual discrimination, specifically, appears problematic. Fourth, we'll address some independent objections to our proposed account.

Start with an observation. Paradigmatic cases of discrimination—such as the *Firing* cases—involve differential treatment that is motivated in a certain way. Boss (gender) discriminates against Employee because they take gender as a consideration *in favor of* firing her. 'Taking gender as a consideration in favor of" can be understood as picking out a *motivating reason* that explains why

[9] There's an interesting question here that we're sidestepping: Do *all* cases of wrongful discrimination involve vicarious wrongs? If so, we would need to know how to delineate the vicarious victim in cases like *Firing #2* where Boss responds to gender. Would the vicarious victim here be Employee qua being a female, or all currently existing females, or perhaps even 'womanhood' as such (if we can make sense of this as a victim with moral standing)? We thank an anonymous reviewer for discussion of this point.

Boss acted as they did (Eidelson 2015; 2022). This motivating reason seems to do crucial work in the vignette. If this detail was absent from the vignette, we couldn't obviously determine it to be a case of gender discrimination, conceptually speaking.[10] Moreover, it would be hard to intuit that something had gone wrong in the first place absent this detail. So, there are compelling reasons to focus our attention here.

But is there any normative flaw we are entitled to infer the presence of from the admittedly sparse information we get in the Firing-type cases? Here's a possibility. Let's say that an agent, A, succeeds in being *substantively rational* just in case A is responsive to the reasons that are available to them.[11] Moreover, let's understand 'having reasons available' in the evidence-relative sense to the effect that the reasons you have available to you are a function of the evidence that is available to you (see, e.g., Fogal and Worsnip 2021; Kiesewetter 2017; Lord 2017; 2018). Finally, let's assume that substantive rationality reflects a normative requirement to the effect that A is subject to criticism for failing to be rational in the substantive sense (see, e.g., Kiesewetter 2017: ch. 2).[12]

Substantive rationality failures are commonplace. To illustrate, suppose that Mark has overwhelming evidence available to him suggesting that Covid-19 vaccines are effective. And yet, he ends up forming the belief that they're in fact not. Intuitively, Mark is subject to the charge of being irrational; he isn't being appropriately responsive to the reasons that are available to him, here coming in the form of evidence for the belief that COVID-19 vaccines are effective. In these cases, the irrationality involves responding inappropriately to epistemic reasons. But parallel forms of irrationality are possible when failing to respond properly to other forms of reasons, such as practical or moral reasons.

[10] See Lippert-Rasmussen (2013) and Eidelson (2015) for conceptual analyses supporting this point.

[11] We understand this broadly to also require *not* being responsive to reasons that lack reason-giving force.

[12] Admittedly, the claim that rationality is normative is not uncontroversial (Broome 2007; Kolodny 2005). But we are happy with making the assumption because even sceptics about the normativity of rationality do not deny that "ordinary attributions of irrationality are commonly understood as criticism … [T]his means that those who deny the conclusion of the criticism argument are committed to a quite radical error theory about ordinary irrationality ascriptions. They are committed to holding that we are all deeply misled in believing that there is anything criticizable about irrationality. And they owe us an explanation of why we all went wrong in thinking this" (Kiesewetter 2017: 39, 41). So if you want to deny this assumption to challenge our view, and thus our suggestion of the flaw in consensual discrimination, you take on a heavy burden of proof.

Before we proceed, let's offer two negative observations to sharpen our understanding of substantive irrationality. First, you can be substantively irrational even if your action or attitude is consistent with what's required by the relevant normative reasons. To illustrate, suppose that you correctly believe that the vaccines are effective but that you do so because of wishful thinking, not because of the evidence available to you. Even though you have a true belief, you intuitively failed in being appropriately reasons-responsive; your belief was improperly based. Second, you can be rational in the substantive sense and yet get things wrong. To illustrate, suppose that COVID-19 vaccines are in fact ineffective, but the normally trustworthy government has succeeded in suppressing all evidence pointing in that direction and ensured that all the available evidence looks credible and suggests that the vaccines are effective. In this case, you're plausibly *not* substantively irrational if you believe that vaccines are effective from the available evidence.

Suppose there exists such a thing as substantive irrationality. Why think that it's typically, or perhaps even necessarily, found in paradigmatic cases of discrimination? One reason for thinking so is that many would seem to agree that treating people differentially based on facts such as their gender, race, or religion (i.e., protected traits) amounts to treating them based on—for most purposes—*irrelevant* considerations. Let's call the claim that, for most purposes, people's race, gender or religion are irrelevant considerations *protected trait neutrality*.[13] We're not going to offer a comprehensive defense of protected trait neutrality given the scope of this paper, but we want to say three things in its defense. First, few, if any, take the challenge of explaining what makes paradigmatic forms of discrimination—such as gender or race discrimination—objectionable to be the challenge of explaining what makes these forms of treatment objectionable *in spite of* the fact that facts about gender or race have reason-giving force. This suggests that most assume that these

[13] Protected trait neutrality is consistent with there being cases where protected traits are reason-giving. For example, it is consistent with saying that gender is relevant when hiring an actor to play a historical female figure. Our sense is that in most (if not all such cases), a protected trait will have genuine, but derivative, significance because of an underlying practical norm conferring significance on the protected trait. For example, in the context of producing a realistic movie, it is desirable to have a lead actor that corresponds to the historical figure depicted on key visual characteristics, such as for instance gender.

facts lack reason-giving force in many of the contexts where we're concerned about discrimination. Second, many have in fact suggested that there is a close connection between paradigmatic forms of discrimination and treating people based on irrelevant considerations. As an example, Lippert-Rasmussen seriously considers, but ultimately rejects, that a defining feature of discrimination is that it's based on irrelevant considerations (2013: 23). Third, taken at face value, it's indeed hard to see how facts about gender or race could have much normative significance in and of themselves, either from the perspective of moral norms or practical reasoning more generally. To illustrate this point, it feels natural to judge that Boss's brute preference for having males rather than females employed is indeed rationally suspicious.

If *protected trait neutrality* holds true, it would naturally follow that discriminatory treatment in Firing-type cases involves substantive irrationality. If you take a fact as a consideration in favor of something that (the evidence available to you suggests) in fact lacks reason-giving force—such as a person's gender—then you fail to be appropriately responsive to the reasons available to you. This idea suggests a rather straightforward analysis of *Firing #2*: Boss's conduct is normatively defective—despite being consented to—because it's substantively irrational in virtue of responding to a fact—gender—that lacks reason-giving force. We can generalize this idea:

> *Irrationality Account*: Discrimination is characteristically objectionable when, and because, it involves substantive irrationality.

We are not the first to float the idea that wrongful discrimination characteristically involves a breach of rationality (e.g., Alexander 1992; Schauer 2003; Lippert-Rasmussen 2013 for discussion). So the contribution we are offering here involves at most a reasonably precise statement of the idea (suggesting the *kind of* rationality that is plausibly at stake) as well as showing how a version of this thought is explanatorily well-positioned to make sense of cases of consensual discrimination. Let's turn to this latter point.

Here are two reasons why the inability of consent to make discrimination normatively appropriate naturally supports a view like the *Irrationality Account*. The first is simple: The characteristic function of consent is to eliminate normative reasons that would otherwise speak against *X-ing*. To illustrate this point, if Moe issues his consent to Lisa borrowing his car, Moe removes a normative reason that would speak against Lisa using his car.[14] Notice now that the *Irrationality Account* doesn't pick out a particular kind of reason against some activity. Instead, its diagnosis is that Boss is guilty of being motivated by a fact that is void of reason-giving force; this fact doesn't speak in favor of firing Employee, nor does it speak against it. Since the flaw here doesn't involve acting contrary to a normative reason that would tell against firing Employee, the flaw isn't the kind of normative property that consent could possibly affect. This reveals at least one deeper, systematic connection between the insignificance of consent to the normative status of discrimination and substantive irrationality.

Here is the second, and slightly more complicated, story. Suppose we're wrong in saying that the problem is Boss being motivated by a fact that lacks reason-giving force. The natural alternative hypothesis would be that Boss's conduct is problematic because there's a normative reason that speaks against firing Employee on the basis of her gender.[15] That result would seem problematic because of the implausible moral unfreedom this would create which would rule out Boss and Employee's interaction.[16] To illustrate this, suppose that Boss and Employee agreed that it would make most sense for both to part ways, and neither is being motivated by considerations of gender (let's call this case *Mutual Separation*). Intuitively, this would seem unproblematic. But presumably, if Employee's gender constitutes or tracks a normative reason against ending the professional relationship in *Firing #2*, then we should expect such a normative reason to exist as well in *Mutual Separation*. (This is plausible because normative reasons exist independently of motivating

[14] By contrast, consent doesn't by itself give Lisa a normative reason to use it. (Suppose, for instance, that Lisa has no need for a car; it would be inappropriate for her to take Moe's consent as a reason for using it).

[15] We take this to be the natural alternative hypothesis here, since if gender was a fact counting *in favor of* firing Employee, it's very hard to see why Boss's conduct would be problematic.

[16] See Tadros (2016: ch. 3; 2020: ch. 4) for extended discussion of appealing to moral freedom in an argument.

reasons, and all that have changed between *Firing #2* and *Mutual Separation* is the motivation of Employee and Boss). That result seems problematic in two ways. First, it just seems counterintuitive to say that *Mutual Separation* is problematic because Boss and Employee fail to recognize that Employee's gender constitutes or tracks a reason against ending their relationship. This suggests that no such normative reason exists. Second, were such a reason to exist, it would go against the interests of both Boss and Employee, since it would imply that they couldn't be fully justified in ending their professional relationship.[17] This would unduly diminish their freedom and control over their own lives and relationships for the sake of no obvious benefit.

To be sure, this second reason doesn't uniquely support the *Irrationality Account* as it only shows that an alternative strategy for making sense of cases of consensual discrimination comes with significant costs. And yet, it supports its core rationale according to which the problem in cases of consensual discrimination is acting on a consideration that lacks reason-giving force. This is because this feature of the view ensures that the *Irrationality Account* doesn't deliver implausible verdicts in cases like *Mutual Separation.* If the problem in Firing #2 is that Boss acts on a consideration that lacks reason-giving force, there is no reason to think that gender supplies a normative reason against mutual separation.

In sum, we‘ve argued that consensual discrimination poses a puzzle and offered a possible solution that follows rather naturally when reflecting upon the nature of consent. We recognize, though, that there's much more to say to get the *Irrationality Account* off the ground. As this task is beyond the scope of this paper, let's instead respond to some worries that may be pressed against it.

5. Objections to the Irrationality Account

[17] This presumes that this normative reason isn't itself sensitive to Employee's consent. One could of course deny this to avoid the challenge we are pressing. But that would imply denying what strikes us as independently plausible and what we've been taking for granted in this paper: There's some flaw in cases of discrimination whose normative status is not sensitive to consent.

A first worry that one may have is that the *Irrationality Account* is incomplete. Call this the *Incompleteness Objection*. Why press this worry? One reason is that if we're right that taking people's gender as a reason for treating them in a certain way amounts to a substantive form of irrationality, then there must be a deeper account to be had of what the relevant considerations and values are in the relevant context (e.g., the context of employment). This account would predict that, and explain why, gender is insignificant. And, the objection could continue, this account is what ultimately explains what makes discrimination objectionable in cases of consensual discrimination.

Here is our two-fold response to this concern. First, we concede that if our account is on the right track, then there might indeed be such a deeper account to be had that would explain why it is that gender is irrelevant in cases like *Firing #2*. And we're also happy to concede that it might be valuable to have this account. What we want to push back on is the thought that having this account would necessarily contribute much to explaining what the characteristic flaw of *discrimination* amounts to. We think that the challenge of explaining or justifying the normative insignificance of gender is a different task from explaining what makes discrimination (for instance, based on gender) characteristically objectionable. And it seems to us that the *Incompleteness Objection* conflates these two questions. Second, in all likelihood, what practical considerations that are worth taking into account vary depending on details of the context (for instance, making decisions about hiring and firing may involve other aims and values than making decisions about, say, delivering public services). But it's natural to think that discrimination bears a characteristic context-invariant flaw in each of these contexts. And the kind of context-dependent account asked for to answer the *Incompleteness Objection* would seem ill-equipped to explain this. That is to say, we think there's useful generalization about the normative profile of discrimination to be had that presupposes rejecting the perspective suggested by the *Incompleteness Objection*.

A second worry is that we've been developing our argument with reference to a perpetrator with a very specific motivational profile. But it's common to recognize that you can wrongfully discriminate via different motivational profiles. For example, it's possible to discriminate based on implicit bias and it's possible to indirectly discriminate. Is the account capable of generalizing to such cases?

Start with the case of implicit bias (gender) discrimination. Such cases have the feature that the discriminator responds to the gender of a victim (or their perception of the gender of the victim) in a way that leads to differential treatment, but where it doesn't make sense to trace this response via the conscious mental states of the discriminator. Instead, we need to appeal to functional (causally operative) mental states to explain how it is that the gender of the discriminatee prompts disadvantageous treatment. Our account can explain what goes wrong in such cases, provided that we don't interpret the motivating reasons-clause too narrowly here.[18] Plausibly, it seems possible to respond to considerations implicitly: I might decide spontaneously to jump into the shallow pond to rescue a drowning child without giving it explicit thought. In this case, it seems plausible to say that I've succeeded in letting myself be motivated by something that tracks the right normative reasons—that I have a dispositional sensibility that tracks the relevant reasons, as it were. On this picture, it's possible to explain what goes wrong in cases of gender discrimination where people respond to implicit gender biases: the dispositional sensibility manifested in these implicit gender biases doesn't track the relevant reasons, namely that gender has no normative relevance. (Of course, it's hard, perhaps impossible, to imagine direct equivalents of *Firing #2* in the case of implicit bias, because Boss's sensitivity to consent and gender interacts in a peculiar way. But that doesn't undermine the principled point that our account can explain what goes wrong in cases of discrimination based on implicit bias).

What about cases of indirect discrimination? To make this concrete, imagine that a company includes a height requirement in their employment policy because working there requires operating a machine that you can only operate if you're of a certain height. This would be a paradigmatic example of indirect discrimination because it involves selecting candidates based on a

[18] For such an account, see Johnson (2019). For this suggestion specifically in the context of discrimination, see Eidelson (2015).

"facially neutral" rule that has a disparate impact. This disparate impact ensues because gender and height are non-accidentally correlated (females are, on average, lower than males). What the *Irrationality Account* will say here will depend on the finer details. But one way to describe the case is that the company responds to a genuine reason (height is relevant for operating the machines) and so isn't substantively irrational.[19] If we assume this motivational story, then the *Irrationality Account* would deliver the verdict—plausibly, we think—that the company doesn't manifest the kind of flaw we've diagnosed in cases like *Firing #2*.

More generally, many believe that when indirect discrimination goes wrong, this is because the facially neutral rule that's being applied has a *disproportionate* impact on certain groups (e.g., Lippert-Rasmussen, 2013). And, admittedly, some such cases would involve substantive irrationality. To see this, imagine that the height requirement, despite being *pro tanto* justified in light of the company's operations, is wrongful because it has a disproportionate impact on females. In other words, the costs that the rule imposes on females (because of their increased propensity for being denied a job) cannot be justified with reference to the benefits for the employer (and the employees) in having employees that have the height needed to operate the machines. If we assume that this reason was available to the company, then they would indeed fail to be substantively rational because they would fail to be appropriately reasons-responsive. But even still, there seems to be a meaningful distinction to be made between this case and cases like *Firing #2*: In *Firing #2*, Boss is motivated by a fact that is void of reason-giving force, whereas when indirect discrimination goes wrong and substantive irrationality is involved, this is because someone has ignored or overlooked a reason (i.e., a disproportionate balance between costs and benefits).

A third challenge to our account is that the normative flaw it picks out may be thought to be too insubstantial. Intuitively, discrimination is typically taken to be a grave form of wrongdoing, and it's not necessarily a grave form of wrongdoing to be substantively irrational. To illustrate this,

[19] They could also be using the rule because they know that it has a disparate impact on gender, and this is what they want to achieve. This is what Eidelson (2015: 41) calls *second-order discrimination*. This would bring this case in line with *Firing #2* from the perspective of the *Irrationality Account*.

recall the COVID-19 example discussed earlier. It's not plausible that this case involves a grave form of wrongdoing, even if the charge of irrationality is apt. We have two responses to this worry. The first is to concede that typical and paradigmatic forms of discrimination may well amount to grave forms of wrongdoing. The relevant test here is whether consensual discrimination should necessarily be considered a grave form of wrongdoing. And that strikes us as far from obvious. Moreover, a plausible explanation of this would be that (Employee's) consent succeeds in purging the discriminatory treatment of many of the grave wrong-making factors we tend to associate with discrimination (as we've seen in our discussion of the prominent wrongness accounts). Hence, when we take into consideration the explanatory challenge that consensual discrimination poses, we deny that it would be a desideratum of an explanation that it picks out a particularly weighty form of wrongdoing. The alternative view we propose is that we keep the structure of the characteristic normative flaw of discrimination fixed and let it be the contingent features of the context that determine the gravity of this flaw.

A second response is that our account isn't committed to saying that substantive irrationality is everything that goes wrong in typical cases of discrimination. We're only saying that it's *a wrong* that is characteristic of discrimination. Nowadays, many say we should be *pluralists* about what determines the normative status of discriminatory acts (e.g., Moreau 2020; Lippert-Rasmussen 2023). On such a view, we could simply say that perhaps, upon closer inspection, it turns out that it isn't the fact that an act is 'discriminatory' that matters the most to assessing its normative status (here we take 'discriminatory' to mean something that in part involves acting from considerations such as people's gender, race, and so on). Much more important, on this picture, are the things that contingently come with discriminatory acts, such as harmfulness, disrespect or inequality. Of course, you may reject pluralism, but we'd be inclined to think that this leaves you with a steep challenge: Identify some feature in virtue of which consensual discrimination is normatively deficient that *also* explains the weight of the wrong we tend to ascribe to (non-consensual) forms of discrimination.

6. Conclusion

Space won't allow us to develop the *Irrationality Account* here in greater detail. But we take ourselves to have shown how consensual discrimination provides a so far unnoticed and critical case for testing the plausibility of an account of what's bad about discrimination. We've shown that popular accounts struggle with accommodating this case, and we've indicated the starting points for a more promising account—an account, the *Irrationality Account*, which says that discrimination is characteristically objectionable when, and because, it involves substantive irrationality.

You may not be fully convinced by the *Irrationality Account*. But, as we have seen, prominent accounts of the wrongness of discrimination cannot explain why there's any deficiency in cases of consensual discrimination. So in case you don't want to endorse the *Irrationality Account*, you're left with three options: reject the intuitive verdict that there's any deficiency in cases of consensual discrimination, provide a new account of the wrongness of discrimination, or rehabilitate one of the accounts we've suggested cannot accommodate consensual discrimination.

References

Broome, John. "Wide or Narrow Scope?" *Mind* 116, 462 (2007): 359-370.

Dougherty, Tom. *The Scope of Consent*. Oxford: Oxford University Press, 2021.

Eidelson, Benjamin. *Discrimination and Disrespect*. Oxford, New York: Oxford University Press, 2015.

Eidelson, Benjamin. "Dimensional Disparate Treatment." *Southern California Law Review* 95, 4 (2022): 785-855.

Enoch, David. "Giving Practical Reasons." *Philosophers' Imprint* 11, 4 (2011): 1-22.

Enoch, David. "False Consciousness for Liberals, Part I: Consent, Autonomy, and Adaptive Preferences." *The Philosophical Review* 129, 2 (2020): 159-210.

Fogal, Daniel and Alex Worsnip. "Which Reasons? Which Rationality?" *Ergo* 8, 11 (2021): 306-343.

Grill, Kalle. "Antipaternalism as a Filter on Reasons." In *New Perspectives on Paternalism and Health Care*, edited by Thomas Schramme. Springer, 2015: 47-63.

Groll, Daniel. "Paternalism, Respect, and the Will." *Ethics* 122, 4 (2012): 692-720.

Hurd, Heidi M. "The Moral Magic of Consent." *Legal Theory* 2, 2 (1996): 121–146.

Johnson, Gabbrielle M. "The Structure of Bias." *Mind* 129, 516 (2020): 1193–1236.

Jonker, Julian. "Beyond the Comparative Test for Discrimination." *Analysis* 78, 3 (2019): 523–536.

Jonker, Julian. "Directed Duties and Moral Repair." *Philosophers' Imprint* 20, 23 (2020): 1-32.

Kiesewetter, Benjamin. *The Normativity of Rationality*. Oxford: Oxford University Press, 2017.

Kolodny, Niko. "Why Be Rational?" *Mind* 114, 455 (2005): 509-563.

Lippert-Rasmussen, Kasper. *Born Free and Equal: A Philosophical Inquiry into the Nature of Discrimination*. Oxford: Oxford University Press, 2013.

Lippert-Rasmussen, Kasper. *Luck Egalitarianism*. London: Bloomsbury Academic, 2015.

Lippert-Rasmussen, Kasper. "Is discrimination wrong because it is undeserved?" *Inquiry* online first (2023): https://doi.org/10.1080/0020174X.2023.2186947

Lord, Errol. "What You're Rationally Required to Do and What You Ought to Do." *Mind* 126, 504 (2017): 1109-1154.

Lord, Errol. *The Importance of Being Rational*. Oxford: Oxford University Press, 2018.

Moreau, Sophia. *Faces of Inequality: A Theory of Wrongful Discrimination*. Oxford University Press, 2020.

Nozick, Robert. *Anarchy, State, and Utopia*. New York: Basic Books, 1974.

Schaber, Peter. "The Volenti Maxim." *The Journal of Ethics* 24 (2020): 79-89.

Tadros, Victor. "Consent to Harm." *Current Legal Problems* 64, 1 (2011): 23–49.

Tadros, Victor. *Wrongs and Crimes*. Oxford: Oxford University Press, 2016.

Tadros, Victor. *To Do, to Die, to Reason Why: Individual Ethics in War*. Oxford University Press, 2020.

Thomsen, Frej Klem. “No Disrespect - but That Account Does Not Explain the Badness of Discrimination.” *Journal of Ethics and Social Philosophy* 23, no. 3 (2022): 420–447.