Defensiveness and Identity
Audrey Yap and Jonathan Jenkins Ichikawa

Draft of April 25, 2022
Forthcoming in Journal of the American Philosophy Association
Draft manuscript. Please cite published version.

Abstract: Criticism can sometimes provoke defensive reactions, particularly when it implicates identities people hold dear. For instance, feminists told they are upholding rape culture might become angry or upset, since the criticism conflicts with an identity that is important to them. These kinds of defensive reactions are a primary focus of this paper. What is it to be defensive in this way, and why do some kinds of criticism, or implied criticism, tend to provoke this kind of response? What are the connections between defensiveness, identity, and active ignorance? What are the social, political, and epistemic consequences of the tendency to defensiveness? Are there ways to improve the situation?

§1. On Defensiveness
You’re not perfect. This, we hope, isn’t surprising or threatening news. We’re not perfect either: we’re not perfectly virtuous or rational, or perfect academics, or perfect anti-oppression activists. And we hope you won’t take it personally if we say: and neither are you. This is technically a criticism, but it’s not a very threatening one. Pointing out imperfection in the abstract typically feels more like a banal truism than an attack.

But sometimes criticism provokes quite different responses — especially when it implicates identities people hold dear. Most people, for example, at least nominally hold anti-racist ideals. If you tell someone that what they just did or said is racist, you are likely to provoke a defensive reaction. They might become angry or upset and refuse to keep talking to you. Such defensive reactions are a primary focus of this paper.

In this paper, we’ll articulate a functionalist characterization of defensiveness, and illustrate harmful ways in which defensiveness can shut down important inquiries. We connect this characterization to the idea of active ignorance, and discuss why some kinds of criticism, or implied criticism, tend to provoke this kind of response. The dynamic is particularly interesting in cases where people with anti-oppressive identities fail to live up to their own ideals. Our central case studies involve "gender-critical" feminism and rape culture; to help us analyze them, we will draw on bell hooks’s discussion of defensiveness and white feminism. We aim to illuminate the social, political, and epistemic consequences of defensiveness, and give some suggestions as to how, depending on one’s social position, one might improve the situation.

1 Co-authoring note: this is an equal collaboration.
We identify defensiveness with the tendency to employ defensive strategies. We'll give some examples of such strategies below, but the characteristic aim of a defensive strategy is to protect someone’s own views or self-conception from perceived or genuine attack.\(^2\)

Here are two important clarifications about our treatment. First, this is a functionalist, as opposed to an affective, approach to defensiveness. We emphasize what one does in a conversation, not how one feels about it. There are, we think, characteristic feelings of defensiveness that may play significant roles in moral psychology. But our project here is more a project in social epistemology: we wish to explore the epistemic effects of defensiveness on conversation and social understanding. The functionalist characterization is the more useful one for this purpose.\(^3\)

Second, our characterization of defensiveness is normatively neutral. One is defensive to the degree to which one tends to employ defensive strategies. While we think defensiveness is often the result of cognitive dissonance, it isn't always. Sometimes, for instance, it is a reaction to accurately perceived unfair attacks. In such cases, it may well be all-things-considered best to take a defensive attitude. Protecting oneself is often appropriate! Yet even when defensiveness is justified, we'll argue, it carries significant social costs.

Our interest in defensiveness is primarily about some of the mechanics and motivations for its production, and the harms that it can create, especially when unwarranted. An interesting project in normative ethics — not ours — would be to demarcate the conditions under which defensiveness is justified. Our contribution is a deeper understanding of the harms of defensiveness, and some strategies for ameliorating them.

\(\S 2\) Defensive Strategies

Defensiveness contrasts with openness. When one is open, one is looking to learn; when one is defensive, one is looking to reject. One way to reject an idea is to articulate arguments against it. But the suite of defensive strategies is much more diverse.

Defensive strategies are often *performative*. We don't mean that they are insincere — just that their expression has strategic use, by centering the interests and feelings of the person being criticized, or possibly being used to forestall additional criticism.\(^4\) In the face of such

---

\(^2\) There are some points of similarity between defensive strategies, as we understand them, and Aaron James’s (2012) theory of assholes: both centrally involve being resistant to criticism. But there are also major differences. James’s assholes, unlike those exhibiting our defensiveness, do not even recognize the humanity or relevance of those who criticize them; defensive strategies are often particularly motivated by the (real or perceived) personal relational nature of the conversation. (p. 27) And James, unlike us, is interested in identifying a stable character trait; real assholes, for James, are assholes in a wide variety of social situations. (p. 20)

\(^3\) We won't set emotional responses aside entirely, but we talk about them primarily as motivators of defensive responses rather than as characterizing defensiveness in the first place.

\(^4\) This is consistent with thinking of such responses as *reactive attitudes*, following PF Strawson (2008, 7–8), as the person who is being defensive is responding to a perceived moral injury. But these issues are gendered
anticipated responses, potential critics might engage in what Kristie Dotson (2014) calls *testimonial smothering*. They might even be (or fear being) pre-emptively attacked, for instance by being accused of hypocrisy.

Here is an example illustrating what we take to be a familiar dynamic. In February 2019, at a US House of Representatives committee meeting, Republican Congressman Mark Meadows used a Black woman as a literal prop in an argument against accusations of racism against then-President Donald Trump. He asked Lynne Patton, a federal employee, to stand up. He did not invite her to speak, instead merely gesturing at her visible Blackness, and spoke on her behalf (denying the possibility of Trump’s racism) as she stood silently behind him.

Democratic Representative Rashida Tlaib criticized Meadows’s action, saying (rightly) that it “in itself it is a racist act.” But her criticism provoked an extreme defensive reaction from Meadows. He complained to Elijah Cummings, the (Black, Democratic) chair of the committee about her harsh words, who focused his energy on placating Meadows: “I can see and I feel your pain, and I don’t think Ms. Tlaib intended to cause you that, that kind of pain.” Tlaib eventually clarified that she did not intend to call Meadows personally racist, and apologized for upsetting him. Meadows’s defensive strategies ended up having a dramatic and disruptive effect on that discourse.

This example illustrates several common defensive strategies: it involves a performative show of anger and woundedness, and re-orientst subsequent discourse around the Meadows’s feelings and standing. These strategies fall neatly under Jennifer Freyd’s concept, “DARVO”. The acronym stands for “Deny, Attack, and Reverse Victim and Offender.” It describes a characteristic way that a subject of criticism may lash out, casting those advancing criticisms as the ones doing wrong.

Meadows’s example also involves another common defensive strategy in the form of *catastrophizing exaggeration*. One might hear a complaint about one’s behaviour, and interpret it as a statement about one’s character — as when the discourse around Meadows focused on whether *he himself* was racist, as opposed to whether his action was a racist one. There is an enticing inference from actions to characters: they did something racist, so they are racist. And another equally slippery inference from there to an even more incendiary idea: they are racist, so they are *a* racist. When we assume that one can chain these inferences together so that doing something racist implies that one is a racist, the stakes are high; nobody wants to be painted with that monstrous brush.

(and racialized in ways that prompt careful contextual treatments of individual cases (see Jaggar 1989, Maclachlan 2010, Hamad 2019, Taiwo 2020 for much more on these aspects).


6 Some — e.g. Anderson (2010, p. 49) — have advocated for restrictive uses of terms like “racist,” due in part its tendency to provoke such reactions. As we will explain in §7, we do not entirely agree, at least in all cases. As Liao & Hansen (2022) carefully argue, “racist” and other such oppression-condemning expressions function linguistically as unexceptional gradable notions, perfectly capable of expressing various degrees or respects in which something is contributing to oppression.
Both of these inferential steps are invalid. While racist people might perform racist actions and hold racist beliefs, so might other people. Most people at least sometimes perform actions or hold beliefs that are out of step with their overall character. Generally patient and considerate people sometimes snap at others for small offenses; and people who are generally committed to being anti-racist might sometimes do racist things out of carelessness or genuine ignorance. Treating criticisms of behavior as attacks on one’s personal character is a common defensive reaction.

Recall that for us, defensiveness is a matter of what one does, not how one feels. Responding to these criticisms in this way isn’t merely a disruptive, unpleasant, and unfair instance of DARVO. It also shuts down conversations and provides obstacles to knowledge — both broader social knowledge, and self-knowledge. And it often does this even when the expressions of frustration are sincere.

That said, we do think there are some typical motivations for defensiveness when it comes to things like accusations of racism. Getting clearer about them can help us understand some of the practical implications of defensiveness in lived reality.

§3. Active Ignorance

We think that many defensive strategies support active ignorance.

To be ignorant about something is to fail to know it. One might be ignorant of something simply because one lacks evidence; you are likely ignorant about our dogs’ names, for instance, if you don’t know us personally. But ignorance can also be active and motivated.

Ignorance fights back when it feels necessary. Sometimes there are things we feel like we need not to know about ourselves in order to maintain our self-conception. What this means is that sometimes having a stable self-understanding can actually require us to be ignorant, either of some facts about ourselves or some facts about the world around us. As José Medina writes,

Social agents can be ignorant in many ways. As we just saw, sometimes there is ignorance out of luxury — when one does not need to know. But sometimes there is also ignorance out of necessity — when one needs not to know. (Medina 2013, p. 34)

Not needing to know is ubiquitous. You might not need to know the bus fare in Victoria because you prefer to drive around Victoria — this needn’t be pernicious ignorance, although it may reflect your economic privilege. Or you might not need to know the bus fare in Victoria for the simple reason that you never spend time in Victoria. What you need to know — and indirectly, what you do know — will depend on your social location in a variety of ways.7

7 This is a starting-point for standpoint theory, but our point here doesn’t require anything as strong as the central elements of standpoint theory, such as the idea that marginalized social locations often put people in
Our main interest here lies, not in not needing to know, but in needing not to know. Here, one’s ignorance is actively maintained, to protect oneself from potentially disruptive knowledge. Sometimes the maintenance of ignorance is the result of oppressive social structures like white supremacy or colonialism. Such ignorance, as Charles Mills (2007, 13) characterizes it, is “an ignorance that resists,” and “fights back.” For instance, in order to maintain the image of Canada as a peaceful settler state, rather than one founded on genocide and dispossession, a great deal of history (and current practice) needs to be concealed (Regan 2010). Settler Canadians who had been miseducated about the truth of residential schools needn’t be operating in bad faith; rather, one can easily form mistaken beliefs because of “social suppression of the pertinent knowledge” (Mills 2007, 21). In such cases, counteracting these practices of active ignorance requires concerted and coordinated social action and educational endeavors, such as the work of the Truth and Reconciliation Commission.

But such work often generates significant friction, in part because individuals often also contribute to the maintenance of their own ignorance. Following Medina, when we speak of ‘needing not to know’, we mean a motivation not to know, in order to protect a particular interest. We do not mean to suggest that it really is best or necessary that settler Canadians not know about the genocidal history of Canada; on the contrary, we think it’s important that all Canadians learn this important history. But they “need” not to know it in order to preserve a particular self-conception they value. It is, after all, very uncomfortable to learn that you are complicit in others’ oppression.8

Gaile Pohlhaus, Jr.’s notion of willful hermeneutical ignorance also helps explain how ignorance is preserved. Pohlhaus focuses on concepts and other hermeneutical resources that have been developed within marginalized communities, but that are actively resisted by members of dominant groups, even when they are introduced to them.9 Racism is a case in point — there is tremendous mainstream discourse about racism’s complex structural power. Anyone willing to listen can easily be disabused of the cartoon picture of racism, where it is perpetuated exclusively by evil people with self-aware discriminatory intentions and beliefs. But many people find it easier to avoid the uncomfortable conversations that disabusing this ignorance would allow. The example of Mark Meadows given above fits this mould perfectly. Tlaib called his action racist, and, in a show of performative woundedness and anger, he refocused the conversation as one about how he himself is not personally racist. (And as a result, the criticism of his racist action was practically forgotten, and he wasn’t held responsible for it.)

There are also broader social epistemic harms. When it is hard for individuals, say, teachers, to gain insight about whether they might be perpetuating racist tropes in their position to achieve superior epistemic capabilities. (We accept that claim too, but we don’t rely on it here.) We here rely only on the truism that knowledge can be impacted by social position.

8 There are many kinds of motivations that can work towards active ignorance. Kinney & Bright (2021) argue that, given certain independently-motivated assumptions about preference and risk, dominantly-situated people may even be rational in preferring not to have further information about their privilege, even if they have no interest in perpetuating oppression.

9 Pohlhaus (2012), following Fricker (2007). See also Frost-Arnold’s (2016, p. 6) defensive ignorance.
classrooms, it will likewise be hard for social groups like schools to gain such insights, thereby perpetuating the ignorance in question. (And of course, in this particular example, there are downstream epistemic and social harms inflicted on students, who absorb those tropes, and are subjected to a less inclusive learning environment.) So there is always a complex interplay between the motivated ignorance of individuals and structural factors that suppress information.

§4. **Marginalization, Identity, and Defensiveness**

Defensiveness will often work alongside other strategies for maintaining active ignorance. One common strategy for maintaining active ignorance is to seek out, or remain inside, an epistemic bubble in which alternative perspectives are excluded, or in an echo chamber, in which alternative perspectives might be heard, but are consistently undermined (Nguyen 2020). Such spaces can feel “safe,” as serious challenges to one’s active ignorance rarely reach through.

But in an increasingly interconnected world, it’s hard to remain in a bubble at all times, even for those motivated by active ignorance. Sometimes a challenge to one’s more comfortable picture of the world makes it through. And even those in echo chambers can know when others take criticisms of their views particularly seriously. It is here that the defensive stance described in §1, and the particular defensive strategies we discussed, are among the primary ways the need for ignorance can be satisfied. Ideological challenges are not to be taken seriously but must instead be refuted or countered.

Defensiveness in general can manifest as an epistemic vice, but there is a more specific vice illustrated in the cases we’re considering. People don’t necessarily exhibit epistemic virtues and vices across the board; while some people might just generally be more closed- or open-minded, it is also common to be closed-minded about some things and open-minded about others. And we are not necessarily good judges of our own tendencies. In what follows, we consider several cases where people’s image of themselves as marginalized, open-minded, and/or progressive exacerbates their defensiveness.

Many systems of oppression are interconnected, which means that sometimes people who are somewhat knowledgeable about one form of oppression may not know as much about it as they think they do, if they don’t understand the ways in which it connects to other forms of oppression. For example, white feminists may know a great deal about gendered oppression, but can still overestimate their knowledge, if they’ve never learned how gendered oppression is experienced differently by women of colour. The kinds of experiences bell hooks (1984) describes in white feminist circles offer paradigm cases of active ignorance. hooks points out that criticisms by black women would go unheard or be dismissed, if they did not echo the views of the white women in the group — and worse, that black women who spoke out would often be branded as angry or aggressive.

Besides the racist association of black people with irrationality and aggressiveness, part of what is going on here is identity-based defensiveness. Here is a telling anecdote from bell hooks:
In 1981, I enrolled in a graduate class on feminist theory where we were given a course reading list that had writings by white women and men, one black man, but no material by or about black, Native American Indian, Hispanic, or Asian women. When I criticized this oversight, white women directed an anger and hostility at me that was so intense I found it difficult to attend the class. When I suggested that the purpose of this collective anger was to create an atmosphere in which it would be psychologically unbearable for me to speak in class discussions or even attend class, I was told that they were not angry. I was the one who was angry.

... Often in situations where white feminists aggressively attacked individual black women, they saw themselves as the ones who were under attack, who were the victims. During a heated discussion with another white female student in a racially mixed women’s group I had organized, I was told that she had heard how I had “wiped out” people in the feminist theory class, that she was afraid of being “wiped out” too (hooks 1984, p. 13).

We note two things about these exchanges. First, they provide clear examples of people who have (correctly) taken on identities as members of oppressed social groups, and — perhaps partly for that reason — failed to recognize their own complicity in related oppression. Second, the strategy invoked is an example of the kind of catastrophizing exaggeration we discussed in §2. hooks initially criticized the syllabus for its omission of all women of colour, and, by the time the story got around, hooks was notorious for “wiping out” her interlocutors in the seminar.

(The “wiped out” vocabulary is extremely zeitgemässig, but contemporary readers may recognize a parallel with the comically lax way in which talk of “cancellation” is thrown around today. If some feminists on social media point out that someone did a microagression, centrist and rightwing discourse may look at it and tut-tut at how everyone gets “cancelled” at the drop of a hat these days. Quite often, of course, they who were “cancelled” weather a few days’ criticism and get on with their careers; two years later, few will remember that they were ever “cancelled”. The same, we are confident, goes for many of the white women that hooks allegedly “wiped out” in that 1981 seminar.)

The white feminists hooks discusses exhibit epistemic vice: their use of defensive strategies derived from a too-rigid attachment to their identity as oppressed. We don’t have to think that these white feminists are epistemically vicious generally to think that they’re being epistemically vicious in this particular instance. There is every reason to expect that the women enrolled in graduate courses on feminist theory in 1981 were relatively open-minded across many fields of inquiry, especially compared to their peers who did not study feminist theory.

Indeed, many of these white feminists had, we are sure, experienced unfair and oppressive attacks on behalf of patriarchy. Challenging gendered oppression is genuinely anti-oppressive work, and such work creates backlash. So many of them would have often been in exactly the kinds of situations that we described in §1 in which defensive strategies are appropriate. They correctly recognized that anti-feminists were attacking them unfairly,
and engaged defensive strategies in the best possible traditions of self-defense. The problem is that, from the inside, reasonable critiques and unfair attacks can feel very similar, and so very naturally tend to produce similar responses. But becoming accustomed to defensiveness can easily lead someone to over-apply these strategies in cases where they are a mistake. True virtue here, as in so many other cases, lies between the two extremes, and can be a life’s work to develop.

So vicious epistemic character in general might not be the best way to understand the resistance to knowledge on display here. Instead, we’ll argue that we can understand the active ignorance of these feminists as a kind of identity maintenance, analogous to the kind of work that needs to be done to maintain an image of a settler colonial state as peaceable and legitimate. (Think of people who say “this is not who we are” after the discovery of yet another colonial atrocity.) We think that this identity maintenance can contribute to the realms of epistemic errors that Medina identifies:

not just areas of epistemic neglect, but areas of an intense but negative cognitive attention, areas of epistemic hiding — experiences, perspectives, or aspects of social life that require an enormous effort to be hidden and ignored. Ignorance in these cases functions as a defense mechanism that is used to preserve privilege.

(Medina 2013, 34)

As in so many cases, recognizing intersectionality is key to understanding the privilege referenced here.

Sometimes unwanted knowledge is too difficult to ignore — for instance, when a classmate or fellow activist persists in calling attention to it. In these cases, one will feel a need for some reason to dismiss it, so it doesn’t interfere with one’s self-conception. This is exactly the kind of reasoning one of hooks’ classmates describes in an apology letter from a fellow student who came to regret her behavior. The white student writes,

In class after a while I noticed myself, that I would always be the one to respond to whatever you said. And usually it was to contradict. Not that the argument was always about racism by any means. But I think the hidden logic was that if I could prove you wrong about one thing, then you might not be right about anything at all. (hooks 1984, 13)

For this student, discrediting hooks served to protect herself from her critiques. Upon coming to terms with her own bad actions and epistemic limitations, hooks’ classmate also found herself at odds with other white colleagues; speaking of a former close friend, she conjectures that “the possibility that we were not the best spokespeople for all women made her fear for her self-worth and for her Ph.D” (hooks 1984, 13). These students’ identities as feminist thinkers and activists were clearly important to them. When those identities were threatened in class by hooks’ arguments, the response was to reject the arguments by discrediting the speaker, rather than reflect on how their conclusions might mandate a change in thinking or behavior. After all, if they accepted criticisms like hooks’ about the exclusion of people of colour from mainstream feminism, that would imply that they had not really been working against gendered oppression generally — only the kinds
of gendered oppression primarily faced by white women. Accepting that you’re not such a great feminist after all, when it is critical to your personal and professional identity, is not an easy thing to do.

Not all identities feel so important. It might be very important to someone that they are a writer, a loving sibling, and a hard worker, but fairly unimportant to them that they are a good guitar player and a resident of Burnaby, even if they believe all of those things about themselves. They might take it completely in their stride if they find out that their guitar playing is actually substandard, but become extremely upset if they find out that their coworkers see them as lazy. (Were they to become defensive about their guitar skills, that might be a clue guitar skills were actually more important to them than they had realized.)

So it is very understandable that many of our paradigm cases of defensiveness are situations in which people face the idea that they are contributing to racism, sexism, or other forms of oppression. For many of us, it’s quite important to see ourselves as basically good people. As we pointed out earlier, when someone’s racist actions are called out, they often react as though they were told that they’re actually a racist monster. This is only amplified when someone is also deeply invested in their identity as someone decent. This is especially true of self-styled activists, who hold anti-oppression identities close to their cores.

§5. Gender Critical Feminism

People who identify as Gender Critical (GC) feminists typically argue that affirming trans people’s (especially trans women’s) gender is a threat to women’s rights. Two frequent but contested claims that they make are 1) that they are (as feminists) concerned with women’s well-being and 2) that they are being silenced for speaking up for women’s rights. The political aims of many Gender Critical people coincide with the political aims of many right-wing groups who seek to restrict the freedom, autonomy, and inclusion of LGBTQ+ people more generally. Yet while it’s sometimes difficult to separate those who support trans-exclusionary politics on ostensibly feminist grounds from those who support it on the basis of straightforward conservatism, it’s important for our purposes to do so here. This is because of the role identity maintenance plays in the defensiveness we’re considering.

Pointing out that trans-exclusionary policies and views are anti-feminist is no challenge to conservative skeptics about trans rights with no affinity for feminism. But the situation is quite different for self-identified GC feminists who argue for trans-exclusionary views on feminist grounds. They need not to know that their exclusionary views and their commitment to feminism are in conflict with each other, exemplifying defensive ignorance. This has clear similarities to the white feminists in hooks’ earlier account, and we’ll return to that after a bit more explanation of what we think is going on.

People’s beliefs and values needn’t be in perfect internal harmony. But sites of tension should merit careful consideration, not denial. For example, plenty of people are environmentalists and also meat eaters. Many of them acknowledge that it might be better for their environmentalism if they didn’t also eat meat. Others might connect their
environmentalism to sustainable farming or hunting practices. But if someone’s devotion to factory farmed meat is supposedly based on environmental grounds, and it’s very important to them that they're both an environmentalist and a factory-farmed-meat-eater, then there might be things they’d need not to know, say about methane production and deforestation, to maintain those identities simultaneously. Defensiveness — maybe accusing their interlocutor of hypocrisy for not being vegan — seems likeliest in this last case, where it becomes a way to sustain a person’s active ignorance.

We're interested in a similar kind of mechanism of active ignorance in those who argue for trans-exclusion on feminist grounds. One form that argument takes requires the background view that the rights of trans women (say, to have access to women's spaces and services) are in conflict with the rights of cis women (say, to use those spaces and services unimpeded). Many feminists have already pointed out that this is a non-problem. For instance, there is often no practical way to maintain trans-exclusionary practices, and gendered spaces have been trans-inclusive for many years in many countries. Since you can't tell whether someone is trans or cis just by looking at them, in order to enforce a policy requiring people to use the facilities matching their birth-assigned sex, you'd need to check whether people are following the rules. Gender non-conforming women (including many cis women) are especially likely to face such challenges, which compromise their ability to use public spaces. So someone might need defensive ignorance to maintain a gender-critical worldview as well as a general commitment to feminism. This means that, even granting the gender-critical framing assumptions, enforcement mechanisms for any such policies at least conflict with the legitimate interests of some cis women. More simply, we agree with Talia Bettcher (2018) and Hay (2020, 10) that GC views are in fundamental tension with feminism.

Like white feminism, the gender-critical worldview best supports the well-being of a narrower class of privileged women — in this case, gender-conforming cis women. As such, both versions of feminism can actively contribute to gendered oppression in some cases. While this might initially seem counterintuitive, recall that there is a structural element to ignorance as well, in which factors that shape one person’s life — for instance, ways that they are oppressed by racism or transphobia — might not be apparent to someone else who is not oppressed in that way. This is how feminism that centres the interests of white women can neglect the ways that non-white women’s oppression can differ. Yet when faced with the criticism that GC views might in fact reinforce gendered oppression, for instance for non-gender-conforming cis women or intersex people, defensiveness is a frequent response.

In many cases, the defensiveness of Gender Critical people takes the form of claims that they’re being silenced, and portrays their speaking up as an act of feminist courage on their part. The writer Julie Bindel describes herself as the target of a campaign of targeted harassment by trans activists, resulting in student protests at her speaking engagements,

10 This happens to gender-non-conforming cis women fairly often. For one high-profile case, see https://www.nytimes.com/2008/05/14/nyregion/14gender.html
11 For much more on the issue of intersectionality and oppression, see Crenshaw (1993).
some of her other events being canceled, and her failure to win a journalism award from the LGBTQ+ organization Stonewall UK. Nevertheless, she writes, she will continue to work as a “tireless, radical feminist” (Bindel 2018). Similarly, after the philosopher Kathleen Stock was awarded an OBE (Order of the British Empire), several academics wrote an open letter critical of that decision, on the basis of her writing about gender. In response, Stock wrote an article in the newspaper *The Spectator*, in which she characterizes that letter as attempting to intimidate and silence those who criticize a prevailing orthodoxy (Stock 2021).

The strategy of catastrophizing exaggeration is apparent here. The open letter critical of Stock explicitly affirmed her academic freedom and her right to express the trans-exclusionary views in question — it simply expressed disagreement and disapproval of those views, and criticized the British government’s decision to award her an OBE. Her response to the letter explicitly paints it as “a document which wouldn't look out of place in the Salem Witch Trial archive” (Stock 2021). And, as in the case of Mark Meadows, the criticism about the effect Stock’s writing has on discourse around trans people became reinterpreted as a claim about her character. (She writes: “The authors of this letter clearly believed they could see into my soul.”)

After Stock resigned her position at the University of Sussex, Bindel wrote an article portraying Stock as the victim of a “witch hunt mentality” that eventually pushed her out of her job (Bindel 2021). Recall Freyd's acronym, DARVO (Deny, Attack, Reverse Victim and Offender) and Frost-Arnold’s defensive ignorance, both describing cases in which someone, when criticized for doing harm, portrays their critic as the “real” wrongdoer. In this case, the LGBTQ+ students at Sussex who protested perceived transphobia are portrayed by Stock and Bindel as aggressive and abusive — as the real offenders. Stock denies being transphobic; whether or not she is correct to do so, the defensive reaction on her part is certainly bound up with her identity as a woman and self-identified feminist. It is precisely because she identifies as both gender critical and a feminist that the defensiveness takes the shape it does, namely accusations that her critics are attempting to silence women for speaking up.

§6. *Rape Culture*

RAINN estimates that slightly over 10% of all American adults have suffered an actual or attempted rape at some point in their lives. Although there is no live controversy as to whether rape is acceptable, feminist theorists have described the world we live in as

---

12 [https://sites.google.com/view/trans-phil-letter/](https://sites.google.com/view/trans-phil-letter/)
13 [https://www.rainn.org/statistics/scope-problem](https://www.rainn.org/statistics/scope-problem), citing the 1998 National Violence Against Women Survey. This includes about ten times as many women as men. In 2019, approximately 3% of all Canadians reported being sexually assaulted, with women about 5 times as likely as men. See Cotter 2019.
embodying a “rape culture” — our society includes norms and practices conducive to the perpetuation of sexual violence and men’s sexual entitlement to women.\textsuperscript{14}

Rape culture is a structural phenomenon, but like all structural phenomena, it is reinforced by individual actions. (Whether an audience is loud is not a feature of, or typically attributable to, any one individual; but individual actions contribute to the broader social fact in obvious ways.) Contributions to rape culture are not always deliberate, but they are harmful. People who personally think that rape is a horrible act might nevertheless contribute to rape culture: they might discount a sexual assault allegation against an acquaintance on the grounds that their interactions with that person had previously always been pleasant, or hypothesize that a woman complaining about being groped was probably inviting the attention, or support policies that make it more difficult for people to raise complaints safely.\textsuperscript{15}

Discussion of rape culture often triggers defensiveness. Despite widespread theoretical discussion of rape culture as a \textit{structural} phenomenon, people who are told that they are contributing to rape culture can easily hear it as an accusation that they are literally in favour of rape. Recall the defensive strategy from §2 in which someone exaggerates a criticism so as to raise the evidential bar.

This tendency toward exaggeration is not always purposeful. Sometimes, it might be the result of a misinterpretation of the label “rape culture”. “X culture” labels sometimes signify a culture where X is literally celebrated — consider e.g. “gun culture”. Someone with limited exposure to feminist thought might be forgiven for thinking personal opposition to rape inconsistent with contributing to rape culture. If they hear that they’re doing so, their likely defensiveness is understandable, but regrettable.

Although defensiveness is a common response when someone is accused of contributing to rape culture, it manifests differently in different cases. Consider generally well-intentioned men. It’s a comforting picture, for men with feminist values, to draw a clean line between the bad men who perpetuate sexual violence, and the good guys (like them) who condemn it and treat the women around them with respect, even serving to protect them from the bad ones.\textsuperscript{16} When a “nice guy” aligned with progressive values is accused of sexual misconduct, defensive reactions (from him or on his behalf) are common. When the

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{14} See e.g. Brownmiller (1975), Burt (1980), Buchwald et al. (2005), Crewe and Ichikawa (2021).
\item \textsuperscript{15} Indeed, as Yap (2017) and Tilton (2022) have emphasized, many common ways of expressing \textit{opposition} to rape actually contribute to rape culture, by perpetuating harmful ‘monster myths’ about rape and rapists. We’ll discuss this dynamic in more detail in §7.
\item \textsuperscript{16} Even this much, while often well-intentioned, arguably reinforces oppressive norms. This idea is closely related to Susan Brownmiller’s insight that the threat of sexual violence is closely connected to men’s ownership over women (1975, 16–17) that some have described as a “male protection racket” (e.g. Card 1991). And even the trope of the “nice guy” (who often ends up in the “friend zone” rather than getting the girl) is itself a contribution to rape culture. This trope is associated with the idea that women often overlook the much-more-deserving nice men around them when looking for romantic partners. But even that framing highlights the background assumption that being a nice guy entitles a person to romantic attention. One of its most extreme cases is exemplified by Elliot Rodger, the Isla Vista Killer, calling himself a “Supreme Gentleman”.
\end{itemize}
\end{footnotesize}
(pseudonymous) "Grace" wrote about the comedian Aziz Ansari pressuring her for sex, she faced vitriolic responses.\textsuperscript{17} It is much easier to align oneself with the #MeToo movement by engaging in strong public criticism of high-profile offenders like Harvey Weinstein than it is to reflect on how one’s own actions might contribute to the problem. Recognizing one’s own contributions to rape culture — whether by victim-blaming, or exhibiting objectification and entitlement to women around them, or by ignoring sexual harassment against women they do not find personally attractive, or in any of myriad other ways — can threaten one’s self-conception as a progressive and decent person. The obstacles that obscure men’s recognition of their contributions to rape culture have also been a common theme amongst feminist theorists and activists.\textsuperscript{18}

But we are especially interested in a different issue, namely the interaction between rape culture, defensiveness, and feminist identities, especially identities attached to feminist women. Self-identifying feminists, even more than other people, have identities that depend for their maintenance on opposition to sexual violence, so we are particularly likely to reach for defensive strategies to resist the idea that we contribute to rape culture. But feminists — including feminist women — do sometimes contribute to rape culture. When they do, their identities as feminists give rise to special challenges.

Here is an example. In November 2016, Canadian feminist author Margaret Atwood signed an open letter criticizing the University of British Columbia for its dismissal of Stephen Galloway, who had been accused by a former student of sexual harassment and assault. The letter’s concern was mainly procedural, but was criticized by several feminist activists for centring Galloway’s interests, and its failure to challenge the status quo of rape culture and an environment that too often silences victims of sexual violence.\textsuperscript{19}

Although she was one of several dozen signatories, Atwood’s feminist reputation made her a particular focus of the controversy. Anti-rape activists who had grown up thinking of her as a feminist icon felt betrayed and disappointed by her position.\textsuperscript{20} We fully acknowledge

\textsuperscript{17} The original account is here: https://babenet.net/2018/01/13/aziz-ansari-28355 and some backlash against her can be found here: https://www.buzzfeednews.com/article/tanyachen/ashleigh-banfield-open-letter-to-aziz-ansaris-accuser

\textsuperscript{18} See e.g. Yap (2017) or Tilton (2022).


\textsuperscript{20} A related phenomenon — outside the scope of this paper — has to do with the role of anti-oppressive identities in making harmful ideas palatable. We think this happens when, for example, one invokes one’s feminist identity as part of one’s efforts to discredit someone’s sexual assault complaint. Kipnis (2017) is, in our view, an example of such a project. Some of the same tropes that would be rightly dismissed as toxic (and tired) contributions to rape culture, coming from dude-bro “Men’s Rights Activists”, were treated as a fresh heterodox perspective, coming from within the feminist camp. (See for example Jill Filipovic’s New York Times review, which compares Kipnis’s book to a similar book by two men. Filipovic rightly recognized the latter’s contributions to rape culture, but wrote favourably about Kipnis’s book, despite its trading in the same one-sided rape culture tropes.) While we don’t see defensiveness as such at play here, this is an instance
and agree with the importance of due process response to sexual misconduct allegations. Our point is not to defend the university's treatment of Galloway, but to point out how Atwood’s response to her critics employed some of the defensive strategies that we outlined above.

One of Atwood’s moves was to exaggerate the criticisms into “the position that the members of a group called ‘women’ are always right and never lie — demonstrably not true — and that members of a group called ‘accused men’ are always guilty.”21 A position like this is a caricature of anti-rape activism; Atwood’s rejection of it is no challenge to the claims her critics actually made. This exaggeration also allowed for a display of performative woundedness. Atwood portrayed herself as a victim of her critics, insinuating that she is being held “responsible for all ills,” for “failing the world on gender equity,” and that perhaps she should “stop trying.”22 She put a person-label front and center in a Globe and Mail op-ed entitled, “Am I a Bad Feminist?” In it, she defends her comparison of UBC’s response to Galloway with the Salem witch trials, and adds comparisons to “the French Revolution, Stalin’s purges in the USSR, the Red Guard period in China, the reign of the Generals in Argentina and the early days of the Iranian Revolution.”

Recall from our discussion of bell hooks’ classmates that unfair patriarchal attacks might establish justifiable and proper defensive habits, which can then be misapplied. Atwood’s case matches the same pattern, we think. As a high-profile feminist author, Atwood has considerable experience as a righteous victim of unfair misogynistic attacks; we have no doubt that, by necessity, she has developed something of a thick skin. But, as in the case of white feminism, it is all too easy for genuine feminist criticism to feel like an unfair attack and a call for defensiveness. Atwood’s op-ed explicitly connects complaints about her defense of Galloway to such anti-feminist backlash:

"It seems that I am a "Bad Feminist." I can add that to the other things I’ve been accused of since 1972, such as climbing to fame up a pyramid of decapitated men’s heads (a leftie journal), of being a dominatrix bent on the subjugation of men (a rightie one, complete with an illustration of me in leather boots and a whip) and of being an awful person who can annihilate – with her magic White Witch powers – anyone critical of her at Toronto dinner tables. I’m so scary! And now, it seems, I am conducting a War on Women, like the misogynistic, rape-enabling Bad Feminist that I am."

As understandable as this error was, Atwood was not listening sympathetically to her critics. If she had been, she would know that none of them said that women are never mistaken or dishonest, or that accused men are always guilty. What they said is that they were disheartened by her prioritization of her literary colleague’s interests over those of the women who had spoken out against him, and her invocation of rape culture tropes in defense of doing so. Portraying her critics in this light had the function of sparing Atwood

of a broader phenomenon, whereby anti-oppressive identities serve to mask one’s contributions to oppression.

21 https://thewalrus.ca/margaret-atwood-on-the-galloway-affair/
22 https://twitter.com/MargaretAtwood/status/952583123157516288
the necessity of questioning the tension between her feminist commitments and her actions. As with the gender-critical feminists discussed in §5, these tensions are sometimes things that one needs not to know about. It is easy to understand how someone could respond to this perceived attack by reaching for defensive strategies, including exaggerating and lashing back against their critics.

§7. Living with Imperfection

Defensive strategies quite generally can give rise to epistemic and social problems. But, as these examples illustrate, there are particular problems that arise when defensiveness serves to protect a person’s strongly held identity. For feminists, a tendency towards defensiveness can sometimes make us epistemically worse off with respect to those very commitments. For instance, as feminists, it is in our interest not to uphold rape culture. And yet, defensiveness can result in our failure to know — or accept — that we may be doing so, when we are. After all, we all fail in our commitments sometimes. We’re also liable to misidentify the cases that do and do not call for defensiveness. None of us are perfect.

Given that we’re all living with imperfection, there are difficult and important conversations that defensiveness might prevent us from having. Many approaches that, in our view, perpetuate oppression — including gender critical rhetoric, descriptions of #MeToo as having gone “too far”, etc. — do so with the stated intention to open space for “important conversations”. These are critical social questions that need to be talked about, their proponents say, in the course of their complaints that less progressive perspectives are being crowded out. We entirely agree that these are important matters that need to be talked about. We need to talk about how to balance due process with supporting survivors. We need to talk about how to satisfy the needs of trans people to exist in society alongside cis people who are uncomfortable with them. But those conversations are impossible without a serious reckoning with the possibility that one is perpetuating harm. hooks’ fellow students were also in such a position, where their focus on their own discomfort meant that they did not consider the possibility that they might be perpetuating oppression.

In §3 we discussed how active ignorance and defensiveness can derail sensitive conversations. Discussion of racism, sexism, transphobia, ableism, and other forms of oppression can all raise the conversational temperature and create hostile and defensive conversations. In all of these cases, serious people — professional philosophers, often — sometimes say that describing actions as problematic in the relevant way (as transphobic, say) is incendiary and unprofessional. Such taboos against criticism represent dramatic

23 We haven’t seen this claim advanced in peer-reviewed publications, but we encounter it often on philosophy blogs and social media — for example in some of the comments to a 2018 blog post on philosophy norms of discourse here: https://dailynous.com/2018/06/02/a-note-on-making-discussions-here-better/, as well as in a blog post about that discussion here: https://theelectricagora.com/2018/06/05/epithets-in-philosophy/. As we mentioned in fn. 5, Elizabeth Anderson (2010, p. 48) articulates a more moderate version of this idea.
obstacles — both political and epistemic. Although we don’t think conversation is always good — we are convinced by arguments (e.g. by Fantl 2018) that sometimes, it’s better not to engage — we do think that productive conversations are quite often an important prerequisite of positive social change.

Being aware that defensiveness can be triggered by the perception of an attack on a closely-held identity might help a potential interlocutor approach a sensitive discussion in a less threatening mode. For example, we think there is often good reason to offer criticism constructively, choose less incendiary words or tones, and focus more on descriptions of actions than negative labels for people. We are interested in producing better conversations and a better culture; less defensive conversations further that goal better than more defensive ones do.

However, people vary in their spaces of options for sensitive conversations. We ourselves — the cis trans-inclusive feminist authors of this paper — have reason to approach cis GC people who are uncomfortable with trans women with at least a degree of empathy and charity, to make a productive conversation more likely. If our interlocutor wants to have a good faith discussion with us about the subject, then we ought to come to the conversation with openness. We ourselves should try not to be unwarrantedly defensive. This will often require listening seriously to criticism from the perspective of ideological stances we do not hold. And if we reject those criticisms, it should be for considered reasons, not a defensive commitment to find fault with the attack. And we certainly shouldn’t respond to good faith engagement with anything along the lines of “how dare you call me a terrible person?” or “why are you trying to get me fired?”.

Of course not everyone shares our privilege. In particular, people whose fundamental rights are challenged by gender-critical stances — trans people, for instance — are being attacked in these conversations in a way that we personally are not. So they may well have good reason to be defensive, to react in anger, etc. Given the bad set of choices available to many victims of oppression, defensiveness and anger may well be fitting responses to attacks on their basic human rights. Such responses can also serve a political end,

---

24 Crerar (2016) describes how taboos can give rise to hermeneutical injustices. We agree. (Crerar’s central example — menstruation — is quite different from ours. We don’t know whether he’d agree with us that many people treat discussion of racism as taboo.)

25 Fantl argues that when one knows that one is right, for instance, open conversation with those who disagree may lead one astray. And closed-minded conversation may be disrespectful or insincere. Some conversations are also just a waste of time.

26 Though as we’ve argued throughout this paper (and seen in our personal lives), even carefully worded criticisms can still trigger defensive responses.

27 Julia Galef (2021) suggests that people tend to react to argument, discussion, or information in either a “soldier mindset” or a “scout mindset”. While we do not wish to rest on any particular empirical commitments of mindset psychology, we do see an affinity between Galef’s soldier mindset and defensiveness, in our sense. A soldier seeks ways to defend their beliefs — must I accept this information, or can I find an excuse for doubting it? A scout seeks ways to correct their misconceptions and augment their knowledge — can I accept this? Are there good grounds for it? Galef says that too much of the time, most people tend to be too much soldier and not enough scout. We agree — especially when it comes to certain kinds of normative criticism.
underscoring the importance of the issue.\textsuperscript{28} As such, we are not attempting here to give a general argument against defensive reactions, or claim that they are never justified. Instead, we simply point out that there is often something epistemically unfortunate about them, even in those cases where they are emotionally proper.

Moreover, as we indicated in our discussions of white feminism and rape culture, the more often one employs defensive strategies, even if they are appropriate, the likelier it is that one will misapply them in cases where they aren’t. We think this is what’s going on when gender-critical cis women justify their defensive stances on the grounds that their rights are similarly being assaulted.\textsuperscript{29} They see trans-inclusion as continuous with the patriarchy, and would likely disagree with being included as examples of undue defensiveness. However, even if one disagrees with the ways we have illustrated the phenomenon, the potential harms of defensiveness still stand. And the normative question of when exactly defensive strategies are justified is not our focus here. We simply explicate a particular kind of harm that often happens when they are employed.

Human relationships are complicated and difficult, and well-intentioned people sometimes make harmful mistakes. That means negotiating accountability is also frequently complicated and difficult. It was natural, for example, for Atwood to want to defend a member of her community. After all, feminists are in a difficult position if we do not believe allegations against our male friends and colleagues, and we do not mean to suggest that there are easy answers. Solutions to rape culture in general are not at all straightforward — and a commitment to anti-oppression does not pull in a single direction. (Anti-Black racism is deeply ingrained in the history of the public imagination of rape.\textsuperscript{30})

But as Atwood’s critics pointed out, we uphold rape culture if we think only of the potential harms to the men accused, and stay silent on the harms to the people who make claims against them. After all, part of dismantling rape culture is the understanding that sexual assault is not the exclusive domain of monstrous people, and a tendency towards defensiveness serves to obscure this in several ways.

First, the defensiveness leads to epistemic costs in the form of false beliefs. Just as the catastrophizing description of a racist calls to mind an extreme exemplar — one that can provide powerful reasons to deny the label to someone who doesn’t match it, even if they are engaging in racist actions — so too does the stereotype of a rapist call up an extreme image of a moral monster.\textsuperscript{31} When the stereotype manifestly doesn’t match the potential

\textsuperscript{28} See, for instance Lorde (1981) or more recently Srinivasan (2018) and Cherry (2021).
\textsuperscript{29} While we acknowledge that they feel that way, feeling it doesn’t make it so. This isn’t the place to argue for it, but our view is that trans-inclusion simply does not threaten cis women’s basic rights the way that trans-exclusion threatens trans people’s.
\textsuperscript{30} Angela Davis (1981, p. 184) describes myths about Black rapists arising after the end of American slavery as a justification for lynchings. See also Hamad (2020, ch. 3) and Tilton (2022).
\textsuperscript{31} Brownlee (2016) argues that using the person-label ‘rapist’ itself wrongs its referents, by essentializing them to their worst actions. We do not go so far as Brownlee in this regard — in at least some contexts, we think it’s important, for instance, for survivors of sexual violence to be able to identify “their rapists” as such — but we agree that there is a moral cost to this language that should not be ignored.
perpetrator, it is easy to conclude, mistakenly, that he must not after all be guilty of the accused conduct: "only a monster would do that."\textsuperscript{32} This is a kind of \textit{modus tollens} response, but a \textit{modus ponens} response is also possible when the evidence of someone’s guilt is overwhelming. If it is only monsters who commit sexual assault, then accepting someone’s guilt will entail accepting that they are a monster.

This epistemic error arguably leads to several social harms. Believing that someone is a monster carries with it ideas about irredeemability, and allows us to write them off, morally speaking. We oppose rape culture; rape is an incredibly harmful act. But we also recognize that many perpetrators of rape are ordinary complex humans with many different life experiences, interests, and impacts on the world around them. Taking rape culture seriously means recognizing the ways in which it desensitizes men to the harms of sexual violence, and portrays domination as desirable. Treating sexual assault as something people do just because they’re monsters lets \textit{culture} — and our shared responsibility for it — off the hook.

Now, what accountability looks like more specifically is complicated, but we want to resist what Elizabeth Bernstein has called ‘carceral feminism.’ Carceral feminism relies on the criminal justice system to help end gendered violence, despite its poor track record at prosecuting sexual assaults.\textsuperscript{33} And we also don’t think that sending offenders to prison (where they themselves face increased risk of sexual assault and other trauma\textsuperscript{34}) is a particularly good solution. Luckily, there is a lot of work out there on alternatives to carceral feminism, for instance in restorative justice, and on feminist prison abolition more generally.\textsuperscript{35}

Nor do we think the kinds of harsh but noncriminal sanctions one sometimes sees — working-class people becoming the Twitter meme of the day and being fired over old racist Facebook posts, for instance — are the way forward. On the contrary, a culture that leaves no room for the toleration of even mild contributions to oppression exacerbates the problem. It further motivates the kind of defensiveness we oppose, and fails to acknowledge that we have all done \textit{something} wrong at some point in our lives.

On top of all this, people change their minds. For example, people who no longer align themselves with GC views sometimes describe coming to realize ways in which their past actions and stances were harmful. Sometimes this is because they see their former allies using unfair defensive strategies. For instance, Beau Dyess became disillusioned with the GC movement when their lesbian friend was attacked — in a textbook DARVO moment — for criticizing prominent GC figures’ homophobia (Dyess 2020). Sometimes such people

\textsuperscript{32} See Yap (2017), Tilton (2022), or Falbo (2022).
\textsuperscript{33} Bernstein (2007). See also Crewe and Ichikawa (2021, pp. 253–4).
\textsuperscript{34} See Alcoff (2018), p. 235.
\textsuperscript{35} See, for instance, Davis et. al (2022) and Kaba (2021) for feminist arguments in favour of prison abolition.
find themselves uniquely well-positioned to have productive conversations with others about these issues.\textsuperscript{36}

Regardless of how we choose which conversations to have, we need to recognize the potential effects of defensive strategies. Shutting down conversations altogether is often a way to stagnate social progress, not to advance it. Defensiveness may often be understandable — or even the best available option — as an emotional response, especially when it comes to already-marginalized people facing unjust attacks. But the use of defensive strategies nevertheless carries significant epistemic costs. There is no simple or guaranteed way to avoid triggering defensive responses in others or to ensure that we are only defending ourselves in ways that are warranted, but being aware of the costs of defensiveness can perhaps lead us to be attentive to the argumentative tactics we use, and the interlocutors with whom we choose to engage.\textsuperscript{37}

\textit{References}


\textsuperscript{36}See for instance Dalwood’s (2021) profile of Alicia Hendley: ““Whatever such people have done in the past, they are in a unique position now to pull others out of the gender-critical movement. That makes them invaluable allies.”

\textsuperscript{37}For helpful discussions on these issues and versions of this paper, we are grateful to Dominic Alford-Duguid, An Dong, Sherwin Arnott, Alisabeth Ayars, Danielle Brown, Marilie Coetsee, Jen Foster, Paul Garofalo, Chris Goto-Jones, Carol Hay, Yong Xin Hui, Carrie Jenkins, Paul Katsafanas, Tim Kenyon, Alec Oakley, Sumeet Patwardhan, Macy Salzberger, Amy Schmitter, Mark Schroeder, Jonathan Weinberg, Scott Woodcock, James Young, three anonymous (to us) referees, and an anonymous (to you) colleague. We presented a version of this paper at the 2021 Western Canadian Philosophy Association meeting in Victoria, BC and the 2022 Conceptual Foundations of Conflict Project Summer Conference; we are grateful to the audiences there for helpful engagement. This paper was supported in part by a SSHRC Insight Grant on Rape Culture and Epistemology.


Cotter, Adam. 2019. “Criminal Victimization in Canada.” Juristat. 113


Goldberg Sanford C. (2011) "If that were true, I’d have heard it by now." In: Goldman Al, Whitcomb D (eds) Social epistemology: essential readings. Oxford University Press, Oxford.

Hamad, Ruby 2019. White Tears/Brown Scars Melbourne University Publishing

Hay, Carol (2020) *Think Like a Feminist*

hooks, bell (1984) *Feminist Theory from Margin to Center*


