Gender and First-Person Authority

Gus Turyn

This is the penultimate draft of a paper forthcoming in *Synthese*; please cite the final version.

1. Introduction

We often think of ourselves as the ultimate authorities over questions about our own minds. When someone asks us how we feel, what we want, or what we plan on doing later, we expect them to take our answers as decisive in virtue of our being in some special position with respect to the answers to those questions. An ordinary way of thinking about this kind of authority is in epistemic terms: we have some kind of special epistemic access to our own minds, so no one else can know better than we do what we are thinking, feeling, or planning.

But views that challenge the accuracy and reliability of our self-knowledge have serious implications for this way of thinking about our authority over our own minds. While we might think of ourselves in everyday life as having excellent—if not perfect—epistemic access to our own minds, philosophers and psychologists have raised significant challenges to this conception of ourselves.[1] We can deceive ourselves; we can be wrong about which stable, long-term attitudes (such as beliefs) we have; we can be wrong about which short-term phenomenal states (such as feelings) we have. That we can be wrong about our own minds in this way raises the question of when, if ever, it is appropriate to challenge another person's sincere avowal that they have a particular mental state. If our authority over our mental states is explained by our epistemic access, then if someone else is in a better epistemic position than we are, they might have better authority over our own minds than we do.

[1] See Williamson (2000) and Schwitzgebel (2006) for arguments against the claim that we have the kind of self-knowledge that we ordinarily take ourselves to have.

Questions about authority can be asked about our gender identities as well: what is the nature of our first-person authority (henceforth FPA) over assertions such as 'I am a man', 'I am a woman', and 'I am nonbinary'? When, if ever, is it appropriate to believe that another person is wrong about their own gender identity? What explains our obligations to treat others in accordance with their avowals about their gender identities?[2]

Talia Mae Bettcher (2009) has distinguished between *epistemic* and *ethical* conceptions of FPA. On the epistemic conception of FPA, our authority over our own gender identities is a matter of our having better epistemic access to our own minds (and thus to our own gender identities) than others do; others' obligations to respect our avowals of our gender identities are explained by our special, first-person access to the contents of our own minds. On the ethical conception of FPA, our authority over our own gender identities is a matter of our having a distinctive kind of autonomy; others' obligations to respect our avowals of our gender identities are explained by their ethical obligations to us as autonomous agents. Bettcher (and others following her) have argued that we should think of our authority over our own gender identities in terms of the ethical conception of FPA rather than the epistemic conception.

While I agree that we have moral obligations to respect the autonomy of other agents and treat them in accordance with what they avow about themselves, I will argue in this paper that an epistemic conception of FPA also should be a desideratum for any trans- and nonbinary-inclusive metaphysics of gender identity. If an epistemic conception of FPA can be included in an account

[2] For the purposes of this paper, I will only consider *sincere* avowals of a person's gender identity. A person can make a sincere avowal even if what they say happens to be false, and a person can make an insincere avowal even if what they avow happens to be true. Cases in which one (knowingly) lies are simply not relevant to my project here; cases in which one is self-deceived or unsure about whether they are sincere, while perhaps more philosophically interesting, are largely beyond the scope of this paper. For the remainder of this paper, I will not include the 'sincere' qualifier. Drawing such a distinction is admittedly more difficult when it comes to forms of transphobia such as those that Bettcher (2007) highlights, such as the view that trans people are 'evil deceivers' intent on misrepresenting themselves to others. However, certain restrictions must be made to keep the discussion here manageable.

of gender identity, then such an account can explain why we should *believe* other people's avowals about their own gender identities rather than merely treat others as if those avowals were true. Developing a metaphysics of gender identity that can explain why we ought to believe others' avowals about their own gender identities has political benefits insofar as, all else being equal, people who sincerely believe what others say about their own gender identities are likelier to live up to the demands of the ethical conception of FPA. Our beliefs govern our actions. Further, disbelieving what others say about their gender identities seems in itself to harm others. An epistemic conception of FPA can better explain how such harms might arise than a purely ethical version of FPA can.

This paper proceeds as follows. In §2, I will discuss ameliorative projects and desiderata currently accepted for accounts of gender identity, focusing on Jenkins's (2018) discussion. In §3, I will introduce Bettcher's account of our authority over our gender identities. In §4, I will discuss several cases which I will argue cannot be accounted for solely by ethical FPA. In particular, I will discuss cases in which others hold paternalistic or condescending views about our gender identities. In §§5-6, I will offer a preliminary diagnosis of why such cases are wrong or harmful and argue that such cases cannot be accounted for under the conception of FPA that Bettcher defends. I will conclude in §7 with discussion of some of the implications of taking on board an epistemic norm of FPA.

My primary aim in this paper is to spell out a problem for thinking about our authority exclusively in terms of the obligations that others have to treat us in accordance with our gender identities: when we merely act as if we believe others about their gender identities despite not believing them, we seem to be doing something wrong, and this does not seem to be accounted for by the dominant understanding of FPA in the literature today. I will not be able to fully solve

this problem in this paper. I will offer a *partial* solution insofar as I will argue that taking this problem seriously should lead us to adopt an epistemic norm of FPA as a desideratum of metaphysical accounts of gender identity, and I will offer the outlines of how we might say that there is something epistemically wrong with disbelieving another person about their gender identity, but I will not do the difficult work of evaluating the prospects of one account or another's ability to incorporate an epistemic norm of FPA. It might turn out that this problem should lead us to abandon the project of spelling out a metaphysics of gender altogether and instead take a deflationary approach such as that advocated for by Louise Antony (2020).

2. Amelioration and FPA

I will approach the question of what kind of norm of FPA we should include in our theories of gender identity through an *ameliorative* lens.[3] Sally Haslanger (2000) has distinguished between a number of ways in which one might approach a question such as 'What is an *X*?'. One might engage in a conceptual inquiry, attempting to spell out the necessary and sufficient conditions for the concept of *X* used in philosophical discourse. One might instead engage in a descriptive inquiry, taking a naturalistic approach and attempting to figure out what (if anything) the folk concept of *X* is. Finally, one might engage in an ameliorative inquiry, which begins by examining the purpose of having a concept of *X* and asking what analysis of *X* would best achieve our ethical and political goals.[4] These kinds of inquiry are not mutually exclusive, as one must have an idea of what the folk and philosophical conceptions of *X* are in order to properly engage in an ameliorative inquiry. But an ameliorative inquiry takes ethical implications to be the central desiderata to be satisfied by a theory of *X*: given the ethical and political work that we want our concept of *X* to do, what should we say that *X* is?

[3] Throughout this paper, I will use 'gender identity' to refer to the property of individuals at issue in discussions such as Jenkins's (2016, 2018).
[4] See Haslanger (2000: 32-34) for discussion.

Different theorists have prioritized different ethical and political desiderata in giving accounts of gender and gender identity. Robin Dembroff (2020), for instance, focuses on genderqueer identities, arguing that the widespread ignorance of such gender identities in analytic metaphysics has given rise to hermeneutical injustice. Elizabeth Barnes (2020) argues that the goal of an account of gender should be to theorize about "what it is—if anything—about the social world that ultimately explains gender" (2020: 3). Louise Antony (2020), in her deflationary account of gender, argues that the key political aim for a theory of gender should be with dispelling the notion that gender differences are in any way natural.[5]

In her (2018) account of gender identity, Jenkins aims towards a concept of gender that can counter transphobia and promote the rights of trans people. Towards this goal, Jenkins offers the following six desiderata on an account of gender identity:

1. The definition should render plausible the idea that gender identity is important and deserves respect.
2. The definition should be compatible with a norm of FPA.
3. The definition should be compatible with the idea that some trans people have a need for transition-related healthcare that is based on their gender identity.
4. The definition should be clear and non-circular.
5. The definition should apply equally well to binary and non-binary identities.
6. The definition should combine well with broader critiques of current gender norms and social structures (Jenkins 2018: 723-724).

As noted above, the goal with which I am primarily concerned in this paper relates to trans and nonbinary gender identities. The higher-order question of which desiderata we ought to center in such debates, while of obvious significance, is beyond the scope of this paper. I will not argue that any theorists concerned with a different ameliorative project are wrong to center the goals that they do; instead, I will merely argue that an epistemic norm of FPA ought to be taken up as a desideratum for theorists concerned with trans and nonbinary identities.

[5] Antony does not explicitly frame her project as ameliorative, but does argue that her account can address political needs (2020: 531).

With respect to the second desideratum listed above, Jenkins leaves open whether we should understand FPA as an epistemic principle or as an ethical principle. If the arguments that I advance in this paper are successful, we can divide Jenkins's second desideratum into two separate desiderata: that the account should be compatible with an ethical norm of FPA *and* that the account should be compatible with an epistemic norm of FPA.

Before I proceed, I want to note that there are two distinct approaches under an ameliorative framework that one can take with respect to FPA. One approach is to focus on gender and gender identity more broadly, accepting as Jenkins does that FPA is a necessary component thereof but nonetheless ameliorating a concept of gender identity that merely takes a need for a norm of FPA as one of its inputs. Call this an *identity-first approach*. A second approach is to focus on the concept of FPA itself and develop a norm of FPA on that basis, then to bring this to bear on the eventual task of ameliorating a concept of gender or gender identity. Call this an *authority-first approach*; this is the approach that I will take here. I will not discuss how other desiderata for an account of gender or gender identity might interact with an account of FPA. Instead, I will treat FPA as a significant phenomenon of its own right and leave the implications of doing so to the side.

3. Ethical and epistemic conceptions of FPA

Let us turn now to Bettcher's distinction between epistemic and ethical conceptions of FPA and to her arguments for the ethical view. Part of Bettcher's aim in doing so is to highlight a form of transphobia that she calls the 'Basic Denial of Authenticity' (2009: 99), which she argues manifests in failing to treat other people's avowals about their own gender identities as authoritative. Bettcher takes FPA to extend beyond gender identity: the denial of authority constitutive of the Basic Denial of Authenticity, Bettcher argues, is the same as the denial of

authority that occurs when one person merely asserts that another person is feeling a particular emotion or when one person insists that another really did consent to sex despite their avowal that they did not (2009: 99; 102; 114). The phenomenon seems to vary in strength across cases; Bettcher notes that a person might permissibly be corrected about whether they really love their sibling, for instance, in a therapeutic context or in a conversation with a very close friend (who, say, notices that their avowals about their sibling do not match their behavior) (2009: 100).

What, then, explains why this phenomenon exists? And what explains its normative force? For Bettcher, it is not a matter of our epistemic access. Bettcher puts the above-mentioned philosophical and psychological pressure on the 'Cartesian' view of self-knowledge in terms of a dilemma for advocates of an epistemic conception of FPA. On one horn, if one attempts to argue *a priori* that we have the kind of self-knowledge that would account for the phenomenon of FPA, their claim is simply false: given the prevalence of wishful thinking, self-deception, and ordinary errors about what attitudes and phenomenal states we have, we simply cannot claim that our self-knowledge is good enough to explain why we should take our self-knowledge to justify the phenomenon of FPA. On the second horn, if one attempts to appeal to the empirical facts about self-knowledge in order to justify FPA, then one will invariably end up weakening the phenomenon as we open the door to the possibility that others are in better epistemic positions with respect to the facts about our gender identities than we are (2009: 101). Regardless of which horn one takes, unacceptable consequences follow.

Rather than conceive of FPA in epistemic terms, then, Bettcher argues that we should think of it in ethical terms. What does it mean for FPA to be an ethical matter? For Bettcher, FPA is intimately related to issues about the privacy of our attitudes and the responsibility we have for the attitudes that we hold. Our attitudes are private insofar as they are not the sorts of things that

others are entitled to pry into or to disclose without our permission; we are responsible for our attitudes insofar as we must 'stake claims' about our attitudes rather than merely treat ourselves as observers gathering evidence about our own attitudes (2009: 101). Violations of another person's FPA, Bettcher argues, are violations of that person's autonomy. As rational agents, we have abilities to determine certain facts about ourselves for ourselves. When another person tries to tell us what our attitudes or our phenomenal states are without asking us or in a way that attempts to override one of our avowals, they try to determine something for us that we have the ability to determine for ourselves. In doing so, they prevent us from making these determinations for ourselves and rob us, at least temporarily, of one of the abilities central to our rational agency.

To illustrate the wrongs or harms constitutive of violations of this ethical conception of FPA, Bettcher appeals to a number of cases. A domestic abuser who believes that his partner has done something to spite him and who coerces a confession from her, even if she did in fact do something to spite him, has wronged her if he takes his own assessment of her mental states to be authoritative. Or suppose that a couple—call them Aaron and Ben—are on a date. If Aaron notices that Ben is tapping his foot impatiently, it would be permissible for him to ask Ben whether Ben wants to go home. But if Ben were to say 'No', then Aaron would be wrong to issue a correction and tell Ben that Ben would, in fact, like to go home. Nor could Aaron simply skip the question and tell Ben 'You would like to go home now'; either response would constitute a violation of Ben's FPA and would be an attempt at control (2009: 102). Both Aaron and the domestic abuser attempt to do something that, if successful, would prevent someone else from exercising an ability that we take to be central to our status as autonomous agents.

We can and should, of course, distinguish between FPA as a phenomenon and FPA as a norm. As noted above, Bettcher builds her ethical account of FPA from the practices present in

resistant, trans-friendly contexts; she does not argue that the same phenomenon exists with respect to gender identity in everyday, ordinary contexts. But Bettcher argues that we can look at the phenomenon as it exists in trans-friendly contexts and use it as a standard that people in ordinary contexts should imitate.

What, then, are the norms that can be derived from an ethical account of FPA? In Bettcher's view, the norms pertain exclusively to behavior. When others avow their attitudes, we ought to treat them as if what they avow is true, regardless of what we might actually believe their attitudes are. When Ben tells Aaron that he's having a good time and wants to remain out on their date, Aaron has an obligation to treat him in accordance with that avowal, regardless of whether he believes that Ben is right about his own desires. Likewise, when someone avows that they are nonbinary, we have an obligation to treat them in accordance with that avowal, regardless of what we believe their gender identities really are. While the latter is more significant insofar as it is more central to one's identity—Bettcher argues that our gender identities constitute an element of our *existential self-identity*[6]—both are matters of the same demand to treat others in accordance with what they avow about themselves.

Jenkins (2018), who follows Bettcher in distinguishing between epistemic and ethical conceptions of FPA, takes the satisfaction of some norm of FPA as a desideratum on accounts of gender identity. For Jenkins, as for Bettcher, the ethical norm of FPA is a behavioral norm: we ought to treat others in a certain way in virtue of their avowals about their gender identities regardless of what we might believe about their gender identities. Jenkins argues that "an ethical norm of FPA is justified by the ethical badness of failing to treat [someone's] avowals as

[6] For Bettcher, existential self-identity is a matter of our answers to questions about who we really are in some deep sense: the kind of persons that we might be even if never given a chance to express ourselves as such (2009: 111).

decisive" (2018: 719) and that such a norm can be adopted even if we have no reason to think that someone is in a privileged position with respect to their own gender identity.

The distinction between epistemic and ethical conceptions of FPA is, I think, quite helpful for understanding the phenomenon of FPA and the underlying features of our authority that we take to explain why the phenomenon exists in the contexts in which it does. But with respect to the norms themselves, the terms 'ethical norm of FPA' and 'epistemic norm of FPA' are somewhat misleading. A norm of FPA that is grounded in the moral obligations that we have to treat others in accordance with their avowals about themselves might include an epistemic component if, as many have argued, our beliefs (and our other attitudes) are appropriate targets of moral evaluation.[7] If we can do wrong by believing certain things, then an adequate ethical norm of FPA will need to account for this. Given this ambiguity in the phrase 'ethical norm of FPA', I will instead distinguish between epistemic norms of FPA and *behavioral* norms of FPA. Epistemic norms of FPA are norms that instruct us to believe certain things (or at least to weigh or acquire evidence in certain ways); behavioral norms of FPA are norms that instruct us to act in certain ways. I will use the phrase 'ethical norm of FPA' to denote any norm that pertains at all to morality, whether action-based or belief-based. If our beliefs are appropriate targets of moral evaluation, then epistemic norms might also be ethical norms.

In what follows, I will argue that a behavioral norm of FPA seems unsatisfactory in light of cases in which one person acts as if they believe another person's avowals about their gender identity despite not actually believing them. In light of the fact that my discussion will proceed under different terms than have been centered in the literature, one might rightly wonder as to

[7] I'll discuss the morality of our beliefs and other attitudes to some extent later in this paper, but I will not take a firm stand one way or another. See Basu (2019a, 2019b) and Basu and Schroeder (2019) for discussion of whether our beliefs themselves can wrong others; see Smith (2005), Sher (2019), and Shoemaker (2011, 2019) for discussion of responsibility for our mental attitudes.

whether epistemic norms are taken to be implicit in the ethical norms standardly adopted—for if they are, then one might worry that I am attacking a position that no one in the literature holds. I think that this worry is misplaced, as Bettcher focuses exclusively on behavioral norms in her (2009) discussion of the issue.[8] Further, even if one thinks that epistemic norms are implicit in the ethical norms put forward in the literature, or even if one thinks that it is obvious that any adequate ethical norm of FPA must include an epistemic component, it is important to make these points explicit. Such a task has not yet been undertaken; I'll attempt to undertake it in the following two sections.

4. The Playing Along Problem

My aim in this section is to begin to apply pressure to a behavioral norm of FPA by considering cases in which one person (at least seemingly) treats another in accordance with their avowed gender identity, but nonetheless seems to deny their authenticity. I will present some cases in which people refrain from believing others about their gender identities and argue that they constitute *prima facie* difficulties for a behavioral norm of FPA.

Begin with the following cases:

> Chris and Dana meet at a bar. Chris is nonbinary. Early in their conversation, Dana misgenders Chris and Chris corrects her. Dana does not believe that anyone lies outside of the gender binary and thus believes that Chris is really a man, but she does not want to offend them and so uses they/them pronouns in reference to Chris for the rest of the evening.
>
> Evan and Fatima meet briefly at a party. Evan overhears Fatima telling another partygoer that she is a trans woman. While Evan takes himself to have progressive views on gender, he holds trans and nonbinary people to higher evidential standards than those to which he holds cis people with respect to their own gender identities. When he believes that someone is cis, he takes their avowals about their gender identities to be decisive evidence and believes them; when he believes that someone is trans or nonbinary, he takes their avowals to be only weak evidence. Nonetheless, he always treats people in accordance with their gender identities so as to avoid any confrontations or accusations of bigotry. As such, he takes Fatima's avowal that she is a woman and her gender

[8] Bettcher has clarified this point in correspondence.

> presentation to be indecisive for her gender identity and suspends belief about what her gender identity is.

At face value, it seems that there is something wrong with Dana's belief about Chris and with Evan's belief about Fatima. Despite the fact that neither Dana nor Evan ever *acts* in a way that is explicitly harmful or wrongful to Chris or Fatima, both would be justified if they felt resentful or hurt upon discovering what Dana and Evan really believed about them.

Why might this put pressure on a behavioral norm of FPA? As discussed above, the behavioral norm of FPA states that we ought to treat others' avowals about their own gender identities as authoritative. And as Jenkins (2018) argues, this norm of FPA is meant to stand independent of any considerations about the evidence that we have for the truth of the claims that others make about their own gender identities: even if we have little reason to believe that others are in the best epistemic positions—or even good epistemic positions—with respect to their own gender identities, we are meant to follow the behavioral norm of FPA and treat their avowals as authoritative.

There is some relevant sense in which both Dana and Evan treat their interlocutors' avowals about their gender identities as authoritative: each defers to what the other says in at least some sense and each takes what the other says to be somewhat action-guiding. In the language of propositional attitudes, both *accept* what the other person says, though neither *believes* what the other person says.[9] Dana and Evan both take on the claim that Chris and Fatima are the genders that they avow themselves to be as premises in their practical reasoning

[9] See Cohen (1989) for some discussions of this distinction. Acceptance is voluntarily acquired; belief is not. Belief in a proposition, for Cohen, is a feeling that the proposition is true; acceptance need not involve any such feeling. We can accept propositions that we do not believe for all sorts of practical purposes—a scientist, for instance, might accept the tenets of their theory even if they do not actually believe them (Van Fraassen 1980). Because both Dana and Evan seem to voluntarily acquire their attitudes and because neither seems to feel that the claims that Chris and Fatima respectively make are true, their attitudes are good candidates for acceptances rather than beliefs.

across all contexts for the duration of their interactions, but to neither does it actually seem to be true; both acquire their attitudes voluntarily and could shed them voluntarily.

Given that both Dana and Evan seem to treat their interlocutors' avowals as authoritative, there is a *prima facie* challenge to the behavioral norm of FPA, and a response must be leveled on its behalf. Should a norm of FPA demand that we actually believe what others avow about their gender identities? If so, does the difference between ethical and epistemic norms collapse? Without answers to these questions, any norm of FPA remains underdescribed.[10]

That this is a *prima facie* challenge to a behavioral norm of FPA does not mean, of course, that it is insurmountable; I will soon consider what a defender of the behavioral norm might say on its behalf. First, however, let me spell out the precise challenge in greater detail. If, as Jenkins (2018) argues, a behavioral norm of FPA is meant to stand independent of any considerations about a person's epistemic standing with respect to their own gender identity—that is, one still ought to treat a person's avowals as authoritative even if they do not take the other person to be in a better or even a good epistemic position with respect to their own gender identity—then some people might still be able to satisfy all that is demanded by an ethical norm of FPA. And if there is something wrong with disbelieving others' avowals of their own gender identities, then the ethical norm of FPA seems to fail to account for everything for which a norm of FPA ought to account.

Call this the *Playing Along Problem*. If a norm of FPA remains silent on whether it is permissible to merely 'play along' with another person's gender identity while at the same time believing that they are mistaken about their gender identity, it seems insufficient.

---

[10] I will emphasize again here that I do not think that this means that the ethical norm of FPA should not be taken as a desideratum on theories of gender identity; instead, I think that it is not enough to explain why all instances of the Basic Denial of Authenticity are wrong.

5. What's wrong with playing along?

The Playing Along Problem is a *prima facie* threat to exclusively behavioral norms of FPA because it suggests that something more is required of us than merely acting as if we believe what others avow about their own gender identities. My aim in this section is to argue that this is indeed a substantive challenge and that there is something genuinely wrong with merely playing along. To make this argument, I will examine a number of potentially wrong-making features that might be present in cases in which one merely plays along.

My aim here is not to argue that any one of these particular features is individually necessary for the Playing Along Problem to be substantive. If any single one of the features that I will discuss below is actually present in cases in which one merely plays along, we have reason to adopt an epistemic norm of FPA as a sui generis desideratum on accounts of gender identity. If, for instance, the only genuine problem with merely playing along is that it is disrespectful and infantilizing, then we have some reason to adopt an epistemic norm of FPA: all else being equal, we should prefer an account of gender identity that identifies loci of disrespect and infantilization and which explains why such instances are wrong. The strength of the epistemic norm of FPA that we should adopt is thus proportional to the number of the wrong-making features that are actually present and to the severity of the wrong in each of them.

With a target now in place, let us begin to examine the potential wrong-making features of Dana's and Evan's beliefs.

1. The probable downstream consequences of their beliefs are different from the probable downstream consequences of other people's beliefs.

What we believe influences how we act. While we might sometimes act in ways that do not conform to what we sincerely believe—when we tell children that Santa Claus exists, when we

try to do things that we know have a very low chance of success, when we tell white lies to preserve our friends' feelings, and so on—our actions are generally guided in part by the way that we take the world to be.

If someone sincerely believes that *p* is true, but thinks that they have pragmatic reasons to act as though *q* is true (where *p* and *q* are two mutually exclusive propositions about the same subject), they are more likely to behave in ways that cohere with the truth of *p* than someone who sincerely believes that *q* is true is. A devoted religious believer is more likely to pray before bed than someone who merely believes that it is in their best interests to act as if they are a devoted religious believer is. A devoted left-winger is more likely to vote for the left-wing candidate than a conservative who has pragmatic reasons to act as if they are a left-winger (say, because they live in a majority left-of-center area) is.

With respect to gender and gender identity, someone who sincerely believes that Chris is nonbinary is more likely to satisfy the demands placed on them by a behavioral norm of FPA than Dana is. Because Dana takes herself to have practical reasons to act as if what Chris avows is true despite not believing what Chris avows, she is more likely to act in a way that violates Chris's FPA. She might do so behind closed doors, as it were—when she thinks that no one will be able to identify that it is her who does so—or when she believes that those present share her beliefs.[11]

If this is right, then we have reason to take an epistemic norm of FPA on board. Requiring that we believe others when they make avowals about their gender identities will increase the probability that we will cohere to the behavioral norm of FPA. While this point does hinge on the

[11] One might object that the behavioral norm only governs our behavior in actual interactions with others, and that situations in which one might be willing to act in accordance with what one truly believes are necessarily not those situations of concern here. But there are a variety of cases, especially those in which we communicate with others in relative anonymity on social media platforms, in which we might interact with others yet feel that we are in enough privacy to reveal our genuine beliefs.

wrongness of our actions, not the wrongness of our beliefs themselves, which norms we take on board as desiderata for a theory should, as discussed above, be informed by the ethical and political ramifications of those norms. And it is clear that there are ethical and political ramifications to what we believe insofar as those beliefs influence our actions.

One might object to this point on the grounds that the downstream consequences of our beliefs are actions, not further beliefs. On this line, one might argue that (potential) actions themselves are what makes Dana's belief problematic: if she managed to cohere perfectly to the behavioral norm of FPA without actually believing what Chris says, then at least along this dimension, there does not seem to be anything wrong with her belief. But as I have emphasized, taking on board an epistemic norm of FPA does not require us to establish that the *belief itself* is what wrongs others. It requires that we establish that there are positive ethical and political upshots of doing so, and those upshots are apparent here.

2. They are committing testimonial injustice.

Failing to believe another person about their gender identity in virtue of their being trans or genderqueer can amount to testimonial injustice. Fricker (2007) argues that testimonial injustice consists in attributing *credibility deficits* to others on the basis of an identity prejudice, where credibility deficits are lower levels of credibility than the evidence would warrant.[12]

Consider how testimonial injustice might occur in the kind of case described above. While Bettcher argues that an avowal of one's gender identity is not merely a report of a fact, reports of facts are involved. We often form beliefs about others' gender identities on the basis of

[12] Others (e.g. Medina 2012, Davis 2016) have pushed back against Fricker's claim that only credibility deficits can result in testimonial injustice, arguing that credibility excesses can as well. While I won't pursue this point in detail here, it seems that the kinds of excess on which Medina and Davis respectively focus can also constitute testimonial injustice in cases relating to one's avowal of one's gender identity. Attributing too high a level of credibility to a trans person's parent or doctor and not to the trans person themself, for instance, can constitute testimonial injustice, as can believing a trans person about a range of subjects beyond what they might reasonably be expected to know about solely because they are trans.

their avowals. When someone tells us that she is a woman, we ordinarily take that as decisive evidence for her being a woman—especially when, for instance, we are not face-to-face with her and cannot see her gender presentation. If we take people's avowals about their gender identities as decisive evidence (or at least strong evidence) for what their gender identities are when they are (presumed to be) cis, then there is good reason to believe that failing to take people's avowals about their gender identities when they are (presumed to be) trans or nonbinary constitutes a kind of credibility deficit. And given the detail of the cases given, it is clear that this is the product of identity prejudices on the part of those such as Evan.

If failing to believe others on the basis of their avowals about what their gender identities are in virtue of taking others to be trans or nonbinary is an instance of testimonial injustice, then we wrong others in failing to believe them about their gender identities. Our wrong is not a matter of the belief that we hold, but instead a matter of the injustice that we do to others in virtue of wronging them as a knower distinctively.

3. They wrong their interlocutors merely by holding an attitude.

There has been significant interest in recent years in questions of whether and how we can wrong others by holding attitudes (such as beliefs) towards them. Much of this interest has focused around questions about the wrongs inherent in holding stereotypical beliefs in general, though a good deal has also been written about how our stereotyping beliefs about other particular individuals can wrong them.[13]

Basu (2019a, 2019b), one of the strongest proponents of the view that our beliefs themselves can wrong others, has focused largely on racist beliefs, but what she says about racist beliefs at least plausibly generalizes to transphobic beliefs.[14] Basu (2019b) has argued in

[13] See Moss (2018) and Munton (2019) for discussion.
[14] See Marušić and White (2018) and Schroeder (2018) for other accounts of how our beliefs can wrong.

particular that even in the absence of any negative downstream consequences, we can wrong others merely by believing racist things about them. Part of Basu's case for this latter point hinges on a thought experiment involving a hermit who, upon finding a picture of someone of Indian descent, forms a racist belief about them. It seems that even though the hermit will never interact with this person, they nonetheless harm them.

We can adapt Basu's case as follows:

> A hermit finds the remnants of a nonbinary person's diary in which they talk about the process of coming to realize that they were nonbinary. The hermit does not believe that anyone is nonbinary and so forms the belief that the person who wrote the diary is actually a man.

It is worth noting that this case seems relatively detached from the actual political concerns facing trans and nonbinary people today; given the violence and discrimination trans and nonbinary people face, there are obviously far more pressing concerns than what those with whom one will never interact think of one. But examining such cases is nonetheless important insofar as it can help us see that beliefs themselves, absent any such violence or discrimination, can still wrong, even if only to a lesser degree than violence, discrimination, and other forms of behavior.

If we follow Basu in holding that individuals can wrong others merely by holding bigoted beliefs about them, then the same might well generalize to Dana and Evan. The seeming badness of what they do can be explained as a matter of their wronging their interlocutors by failing to treat their avowals as epistemically authoritative and forming bigoted beliefs. It is important to note here that this conclusion can still stand regardless of whether one accepts Basu's account of *why* such beliefs are wrong; all that matters here is that we say that such beliefs are somehow wrong in themselves.

4. They act in different ways than people who actually believe their interlocutors act.

Dana's and Evan's actions in using the right pronouns for Chris and Fatima can be characterized under a different range of descriptions than the actions of other people who believe what Chris and Fatima avow about their gender identities can be.[15] If asked why he is using she/her pronouns to refer to Fatima, Evan might sincerely respond 'Because she requested that I use those pronouns to refer to her.' But if pressed further, Evan would reveal a different set of motivations—that he does not want to be confronted or accused of bigotry.[16]

Compare Evan with Gio, who meets Fatima and believes that she is a woman. When asked why he is using she/her pronouns to refer to Fatima, Gio's first response might be the same as Evan's. But upon being pressed further, Gio would reveal a different set of underlying motivations: that he believes that Fatima is a woman and he believes that everyone ought to be referred to with their correct pronouns. In describing both Evan's and Gio's actions, we might say that they are using the correct pronouns to refer to Fatima and that they are doing everything else required at the surface level by a behavioral norm of FPA. But it is also true that Evan is deceiving Fatima and acting solely to further his own interests, while Gio is speaking honestly and acting without ulterior motives.

Given the difference in the range of descriptions that can be applied to Evan's and Gio's respective actions, there is a sense in which they are engaged in different actions.[17] Without taking a stand on the question of whether actions truly can be individuated on the basis of distinct motivations, it is clear that at least certain features of an action—including its moral worth or the degree to which the agent is praiseworthy for it—depend on the facts about what

---

[15] Here, I'll focus on using the correct pronouns for ease of discussion, but what I say about using pronouns generalizes to other forms of behavior as well.

[16] This is, of course, how Evan would *sincerely* respond to such questions. The point here is not what he would actually say if he were asked by another person; the point is that why he acts as he does is different than why another person who sincerely believes that Fatima is a woman would act in what appears to be the same way. We can imagine that this is what Evan would say to himself if he privately questioned his own behavior.

[17] This claim is consistent with Anscombe's (1957, especially §19) influential account of intentional action.

motivated the agent to act. Someone who gives money to those in need because they believe that they have an obligation to do so is more praiseworthy than someone who gives money to those in need because they believe that doing so will impress their friends.

That what we believe about others' gender identities affects the moral worth of our behavior itself becomes more clear when we consider even more egregious examples than that of Evan and Fatima. Take, for instance, the following case:

> Hugh believes that his friend Ignacio, who recently came out as a trans man, is suffering from a delusion. Hugh watches a YouTube video that instructs him to treat Ignacio exactly as he would treat any other man because doing so will help Ignacio realize that he is deluded.[18]

Despite the fact that the words that Hugh utters are the same as the words that anyone who genuinely believes that Ignacio has the gender identity which he avows himself to have would utter, Hugh treats Ignacio differently than other people do. Regardless of whether he succeeds, Hugh is deceiving and attempting to manipulate Ignacio. His actions can and should be evaluated differently than those of someone like Gio's. Hugh is attempting to radically undermine Ignacio's agency and is attempting to manipulate him.

This discussion highlights an ambiguity in how we might interpret a behavioral norm of FPA. On one understanding of the behavioral norm, all that matters is that the right words come out of our mouths: even if we do not believe that others have the gender identities which they avow themselves to have, we can satisfy what is demanded of us by the behavioral norm of FPA by uttering the sequences of words that believers would utter. Should we opt for this interpretation, we can say that others can satisfy the behavioral norm of FPA in the absence of the relevant beliefs and hold that the other considerations discussed here should nonetheless lead us to adopt an epistemic norm as well.

---

[18] Thanks to Joseph Orttung for suggesting this example to me.

On another understanding of the behavioral norm, it does not merely matter that we say the right words; we must say the right words *for the right reasons*. Whether something counts as a particular form of behavior in part depends on the reasons for which we act: whether an utterance counts as a lie, for instance, depends on an agent's background beliefs and intentions. On this interpretation of the behavioral norm, we can say that acting under the range of descriptions that would truly satisfy what is demanded by the behavioral norm of FPA requires that we act in that way because we believe what others say about their gender identities. On this interpretation, it is impossible to satisfy the behavioral norm without satisfying the epistemic norm as well, as the boundary between the two types of norms begins to blur.

There is room to maneuver within both of these interpretations of the behavioral norm of FPA. One might hold that beliefs and intentions matter to a certain degree (e.g. that we ought not to intentionally manipulate others), which would allow us to explain why Hugh does not satisfy what is demanded of him by a behavioral norm of FPA while still allowing that others (like Evan) who take themselves to have progressive views but still happen not to believe what some others say about their gender identities can satisfy the demands.

I will not take a stand here on which of these is the better interpretation of the behavioral norm of FPA. What matters for my purposes is that one can fail to show the right degree of deference to others' authority over their gender identities even if one utters the right words. The behavioral norm, by itself, does not do all that would be needed for a full account of FPA and requires an epistemic norm to go along with it.

6. Objections and replies

There are a number of ways in which a defender of the behavioral norm of FPA might reply to what I have discussed here, and others unsympathetic to this project might, of course,

raise further objections. In this section, I will discuss some potential objections and address them. One might, of course, go systematically through the points that I have raised and object to each of them on distinct grounds. I cannot anticipate every possible objection to the broad range of possible wrong-making features that I have discussed here. Rather than considering replies to individual elements of what I discussed above, I will consider replies that challenge my approach as a whole.

First, one might object that the beliefs on which we ought to focus are not those which I have discussed so far. Rather than focus on Dana's belief about Chris's gender identity itself, one might argue, we should focus on the beliefs which themselves give rise to Dana's beliefs about people's gender identities. After all, believing the right theory about gender or gender identity would likely have the result that Dana would form the right beliefs about Chris's gender identity in turn—for when she is weighing the evidence about Chris's gender identity, one component of that evidence will be what she takes the facts about gender identity to be. If this is correct, then there is some sense in which the solution to this problem will involve an epistemic norm, but the only relevant epistemic norm is that one ought to believe the correct theory of gender identity, which is hardly an interesting claim.

In reply, I will note that we should not ignore the beliefs that individuals such as Dana have about gender in general. But attempting to reduce the wrongness of such cases to matters of an individual's prior beliefs and how those will influence their beliefs about particular individuals' gender identities overlooks the ways in which we can wrong others merely by believing certain things about them. A racist who commits testimonial injustice or who forms a wrongful belief about some other individual ought to alter their beliefs about that particular individual as well as their general beliefs about the subject in question. And as Evan's case

demonstrates, believing the correct account of gender identity is insufficient for forming the right beliefs about others' gender identities. If the sole epistemic norm adopted is that we ought to believe the correct account of gender identity, we cannot account for individuals who believe the correct account but nonetheless fail to believe others' avowals about their gender identities for prejudicial reasons. In sum, while we should aim at correcting others' incorrect views about gender in general, we ought not to think that doing so will completely address the problem at hand.

A second objection that one might raise is that what I have argued for here is too epistemically or metaphysically demanding. A metaphysics of gender identity should obviously be consistent with the strictures of analytic metaphysics and epistemology. If an epistemic norm of FPA is inconsistent with these strictures, then whether or not one accepts my arguments here, one might think that we should not take on an epistemic norm of FPA as a desideratum for theories of gender identity—for doing so would render an adequate metaphysics unreachable. Suppose, for instance, that investigation revealed that the only plausible means of establishing an epistemic norm of FPA would require a 'Cartesian' view of the mind according to which we cannot possibly be mistaken about our own minds. Adopting an epistemic norm of FPA would thus force us to reject the (highly plausible, in my view) arguments against such a view of the mind.

I have two points to make in reply to this objection. First, what I have attempted to establish here is that this is a substantive problem for accounts of gender and gender identity according to which all that a norm of FPA requires us to do is treat others as if what they are saying is true. Whether an epistemic norm is ultimately too demanding to take on board is a matter for further theoretical deliberation, not something that comes to bear on whether it should

be a desideratum on accounts of gender identity. If it ultimately turns out that it is too costly a desideratum to satisfy in the theories that we advance, my project in this paper will not have been undermined.

Second, though a full exploration of this topic is beyond the scope of this paper, I do not think that an epistemic norm of FPA is too theoretically costly. While Bettcher argues that an adequate epistemic norm of FPA requires a 'Cartesian' view of the mind according to which we have infallible, incorrigible, and privileged knowledge of the contents of our own minds, I do not think that all such features are necessary. An epistemic norm of FPA merely requires that our avowals about our gender identities constitute the best evidence available for others to form their beliefs about our gender identities. If our avowals about our own gender identities constitute the best evidence for what our gender identities are, then anyone who does not believe what we say about our gender identities is behaving epistemically irresponsibly: they are believing against the evidence. If we as individuals might be wrong about our own gender identities, but if others could not be in a position to have better evidence about the facts about our gender identities than we have, then others could never be in better epistemic positions than we are with respect to our own gender identities. And our avowals being the best possible evidence for our own gender identities is consistent in principle with, say, a self-identification account of gender identity such as Bettcher's, a norm-relevancy account such as Jenkins's (2018), or a dispositional account such as McKitrick's (2015). Though revisions will have to be made to such accounts in order to explain precisely why they would ground the relevant epistemic claim, such revisions might not turn out to incur too heavy a cost.

Last, one might argue that there is an asymmetry between the cases that I have discussed here and cases in which someone believes that another person is either self-deceived or not

openly identifying as the gender as which they identify. In such cases, one might argue, we might have reason to think that another person actually has a marginalized gender identity despite their public avowals that they do not.[19] And in such cases, it is not clear whether the beliefs that we form about others are wrong in the same ways as are the more explicitly transphobic beliefs discussed above. Granting that there could be cases in which we have reason to believe that a person has a marginalized gender identity but is not open about it, what ought we to do?

Such cases might not be possible if we accept my suggestion that we should take the best evidence for a person's gender identity to be their own avowals—for if we accept this suggestion, then it is not clear how we could be in a position to know that another person is mistaken about their own gender identity. But if such cases are possible, we should clearly defer to the more traditionally construed ethical norm of FPA. If we believe that another person has a marginalized gender identity but is self-deceived or not openly avowing themself to have it, we should not treat them publicly as if they have that gender identity. Instead, we should defer to their public avowals. What I have argued here does not require that we abandon an ethical conception of FPA, nor that we abandon the behavioral norms of FPA. Belief and behavior are both significant and no account of FPA is complete without both.

Nonetheless, this leaves open the question of what we ought to believe in such cases. If we accept the contention that we can wrong others merely by holding an attitude towards them, then we might nonetheless wrong individuals by believing that they are self-deceived about their own gender identities. This is presumably less wrong or harmful than the kinds of cases on which I have primarily focused to this point in light of the fact that the other wrong-making features are less likely to be present (e.g. this does not seem to be a case of testimonial injustice);

[19] Thanks to Kate Manne for raising this objection.

as such, the epistemic norm of FPA for which I have argued applies to cases such as this one as well.

7. Conclusion

How we treat others is obviously enormously significant, but what we believe about others is also enormously significant. This is especially clear with respect to questions about FPA and gender. What we believe about others shapes how we treat others, both to the extent that we are likelier to treat others in the right way if we believe the right things about them and to the extent that we cannot fully treat others in the right way if we do not believe the right things about them. Our beliefs, just like our actions, can be sources of wrongdoing and harm, and an adequate set of norms of FPA must take this relationship into account. Merely playing along with what others have avowed is not enough, epistemically or ethically.

My primary aims in this paper have been to motivate the Playing Along Problem and to argue that taking it seriously requires that we take on an epistemic norm of FPA as a desideratum for accounts of gender identity. While I have briefly touched on some of the epistemological implications of this problem, a full exploration of the epistemology of gender identity—and its resultant metaphysical conclusions—is beyond the scope of this paper. Such a project would involve both an examination of our self-knowledge and our third-personal knowledge of others' social identities—work that lies beyond the scope of this paper. Once it is undertaken, we can reconsider positions in the metaphysics of gender to determine whether they are consistent with an epistemic norm of FPA and adjust them if they are not. As noted above, I am optimistic that extant accounts can at least in principle be revised in such a way that they can be made consistent with an epistemic norm of FPA, but such revisions, too, require attention that lies beyond the scope of this paper.

One might wonder what the practical implications of accepting an epistemic norm of FPA into an account of gender identity might be. We do not have the same kind of immediate control over our beliefs as we have over our actions. How, then, can one actually abide by an epistemic norm of FPA? While a complete explanation of how we can implement this kind of norm into our practices is beyond the scope of this paper, I suspect that the answer lies in cultivating epistemic virtues, as it does with addressing other kinds of epistemic injustice. Rather than attempting to convince others in individual cases that someone *really has* one gender identity or another, we should focus on educating others about gender more broadly in the hopes that doing so will encourage the kinds of epistemic practices that will lead us to believe what others say about their gender identities more quickly when they make avowals about their gender identities to us.

I will conclude here with a brief discussion of how what I have said here about gender might generalize to other social categories. One might wonder whether we should also adopt epistemic norms of FPA with respect to others' races, sexual orientations, and other such categories.[20] I have focused here on gender in part because the literature about norms of FPA largely centers around gender and in part because I hold a broader methodological commitment to treating distinct forms of social identity differently. Our social world is complicated, and what is true of one social category is not necessarily true of another.[21] The considerations that have led us to accept an epistemic norm of FPA about gender identity are not all relevant—though some are—to considerations about race, sexual orientation, or other social categories. I think that the conclusions I have drawn here generalize to sexual orientation; I do not think that they generalize

[20] Thanks to an anonymous referee for encouraging me to say more about this point.
[21] See Barnes and Andler (2020) for discussion of this point.

to race.[22] But a defense of these points would require considerations that lie beyond the scope of this paper and I will not hinge my account on this point one way or another. It might well be that what I have said here generalizes to all other social categories, or it might not generalize to any other social categories. Such discussions are best left for further research.

References

Anscombe, G.E.M. (1957). *Intention*. Harvard University Press.
Antony, Louise (2020). Feminism Without Metaphysics or a Deflationary Account of Gender. *Erkenntnis* 85 (3):529-549.
Barnes, Elizabeth (2020). Gender and Gender Terms. *Noûs* 54 (3):704-730.
Barnes, Elizabeth & Andler, Matthew (2020). *Categories We Live by: The Construction of Sex, Gender, Race, and Other Social Categories*, by Ásta. *Mind* 129 (515):939-947.
Basu, Rima (2019a). What We Epistemically Owe To Each Other. *Philosophical Studies* 176 (4):915–931.
Basu, Rima (2019b). The wrongs of racist beliefs. *Philosophical Studies* 176 (9):2497-2515.
Basu, Rima & Schroeder, Mark (2019). Doxastic Wronging. In Brian Kim & Matthew McGrath (eds.), *Pragmatic Encroachment in Epistemology*. Routledge. pp. 181-205.
Bettcher, Talia Mae (2007). Evil Deceivers and Make-Believers: On Transphobic Violence and the Politics of Illusion. *Hypatia* 22 (3):43-65.
Bettcher, Talia Mae (2009). Trans Identities and First-Person Authority. In Laurie Shrage (ed.), *You've Changed: Sex Reassignment and Personal Identity*. Oxford University Press. pp. 98-120.
Cohen, L. Jonathan (1989). Belief and acceptance. *Mind* 98 (391):367-389.
Davis, Emmalon (2016). Typecasts, Tokens, and Spokespersons: A Case for Credibility Excess as Testimonial Injustice. *Hypatia* 31 (3):485-501.
Dembroff, Robin (2020). Beyond Binary: Genderqueer as Critical Gender Kind. *Philosophers' Imprint* 20 (9):1-23.
Fricker, Miranda (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press.
Haslanger, Sally (2000). Gender and race: (What) are they? (What) do we want them to be? *Noûs* 34 (1):31–55.
Jenkins, Katharine (2016). Amelioration and Inclusion: Gender Identity and the Concept of

[22] I hold this view in part because I am convinced that a social position account of race such as Haslanger's (2000) is right; if an account such as Haslanger's is right, then an individual might not be in the best epistemic position with respect to their racial identity in virtue of the fact that they might not recognize how they are socially positioned. Here, we should take care to distinguish cases in which one individual claims to be a member of one race in order to accrue some social or political benefit (whether a member of a marginalized race who passes as a member of a dominant race or a member of a dominant race who claims membership in a marginalized race in order to take advantage of some opportunity set aside for marginalized individuals) from cases in which a person sincerely believes that they are a member of a different race than they in fact are. In at least some of the former cases, it seems clear that we have good reasons to override others' avowals about their own racial identities, which suggests that we ought not take an epistemic or ethical norm of FPA on board with respect to racial identity. The latter are the more philosophically interesting cases, and I suspect that addressing the normative questions surrounding them would require a good deal more attention than I can provide here.

Woman. *Ethics* 126 (2):394-421.
Jenkins, Katharine (2018). Toward an Account of Gender Identity. *Ergo: An Open Access Journal of Philosophy* 5 (27):713-744.
Marušić, Berislav & White, Stephen (2018). How Can Beliefs Wrong?: A Strawsonian Epistemology. *Philosophical Topics* 46 (1):97-114.
McKitrick, Jennifer (2015). A dispositional account of gender. *Philosophical Studies* 172 (10):2575-2589.
Medina, José (2012). *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and Resistant Imaginations*. Oxford University Press.
Moss, Sarah (2018). *Probabilistic Knowledge*. Oxford University Press.
Munton, Jessie (2019). Beyond accuracy: Epistemic flaws with statistical generalizations. *Philosophical Issues* 29 (1):228-240.
Schroeder, Mark (2018). When Beliefs Wrong. *Philosophical Topics* 46 (1):115-127.
Schwitzgebel, Eric (2006). The unreliability of naive introspection. *Philosophical Review* 117 (2):245-273.
Sher, George (2019). A Wild West of the Mind. *Australasian Journal of Philosophy* 97 (3):483-496.
Shoemaker, David (2011). Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility. *Ethics* 121 (3):602-632.
Shoemaker, David (2019). Hurt Feelings. *Journal of Philosophy* 116 (3):125-148.
Smith, Angela M. (2005). Responsibility for attitudes: Activity and passivity in mental life. *Ethics* 115 (2):236-271.
Van Fraassen, Bas. (1980). *The Scientific Image*. Oxford University Press.
Williamson, Timothy (2000). *Knowledge and its Limits*. Oxford University Press.