E. DÍAZ-LEÓN
Universitat de Barcelona
ediazleon@ub.edu

GENDER IDENTITY, FIRST-PERSON AUTHORITY, AND BELIEF

abstract

Talia Bettcher (2009) has defended a principle known as "The First-Person Authority" (FPA) of gender avowals. *Gender avowals are a specific kind of speech acts, namely, present-tense gender ascriptions from the first-person point of view (which we can also call* gender self-identifications*). This principle has given rise to an interesting debate about what is the best way of articulating and defending the principle in a manner that is useful for promoting the rights of trans people. In this paper, I argue that an adequate explanation of the epistemology of gender self-identifications involves the insight that our knowledge of our attitudinal mental states is based in part on* introspection *of the conscious manifestations of these very dispositional mental states. That is, it is part of the nature of propositional attitudes such as beliefs and desires that they can be* manifested*, under certain manifesting conditions, in the form of* conscious *mental states. Thus, given the strong connection between a disposition and its manifestations, the fact that we can know some of the dispositions' manifestations by introspection explains why I can have an epistemic advantage over other people who are not able to know the conscious manifestations of my mental attitudes by introspection.*

keywords

gender identity, first-person privilege, self-knowledge, amelioration

Phenomenology and Mind, n. 29 - 2025, pp. 114-123
DOI: 10.17454/pam-2910
https://www.rosenbergesellier.it/eng/journals/phenomenology-and-mind

ISSN 2280-7853 (print) - ISSN 2239-4028 (on line)

I. Bettcher's account

Talia Bettcher (2009) has defended a principle known as "The First-Person Authority" (FPA) of gender *avowals*. Gender avowals are a specific kind of speech acts, namely, present-tense gender ascriptions from the first-person point of view (which we can also call *gender self-identifications*). This principle has given rise to an interesting debate about what is the best way of articulating and defending the principle in a manner that is useful for promoting the rights of trans people.

Bettcher (2009) argues that the FPA should be understood as an *ethical* requirement rather than an *epistemic* requirement. Her main argument is that the claim that we have epistemic authority about our gender avowals is strictly speaking false. Her reason for this claim is that we do not have first-person epistemic authority about many of our claims about our own mental lives.

She distinguishes, following other philosophers of mind, between "fleeting" mental states, also known as *occurrent* mental states, and "durable" mental states (also known as *standing* mental states) (2009, p. 99). Occurrent mental states are those that appear in the stream of consciousness at a specific time, have a duration, and typically disappear at a later time (think, for example, of a toothache, an itch, the taste of a hot chocolate, or the smell of a rose). These mental states are always conscious. Standing mental states, on the other hand, do not necessarily enter into the stream of consciousness, and can have a much longer duration. For example, many of our beliefs are standing mental states in this sense: I have had the belief that Paris is in France for many years, and will probably have it for many more, even in those times when I am not consciously thinking about Paris nor France. Standing mental states do not have to be conscious: arguably, they are *dispositional* states that have conscious phenomenal states as some of their manifestations. Desires (and other propositional attitudes such as fears and hopes) are also arguably standing states: for instance, someone can have a long-standing desire to be a University professor. This desire can be manifested in the form of behavior (such as hard work) and other mental states (including conscious states such as feeling happiness and relief when one gets a job offer, etc.).

Bettcher argues, drawing on a consensus in philosophy of mind, that although self-ascriptions of *phenomenal states* seem very close to being *privileged*, we cannot be so certain about our self-ascriptions of *standing mental states*.

Let's examine occurrent mental states first. Indeed, many philosophers of mind reject the Cartesian idea that we have epistemic privilege about our own phenomenal states, that is, they deny that we cannot be mistaken about our own mental states, and that no one can know

about our phenomenal states better than ourselves. This *epistemic privilege* can be understood as the combination of two claims: (1) *infallibility*: if I am in mental state M, I *know* that I am in M; and (2) *incorrigibility*: if I *believe* that I am in M, then it is *true* that I am in M. The second claim is the most important one for our purposes since what is more politically significant is whether first-person gender *judgements* are incorrigible or not, that is, if my belief about my gender self-identification could be wrong, and if someone could correct me about it.

Bettcher argues, in the first place, that with regard to our judgements about our mental states, we are pretty certain about our first-person (present-tense) ascriptions of *phenomenal states*: "avowals of [phenomenal] states generally exhibit an immunity that many judgements do not" (100). Strictly speaking, we *could* be wrong about these judgments: given most accounts of phenomenal states, I could have the belief that I am in pain even if I am not really in that state (because there could be some cognitive mechanism that could be blocking or interfering with the correct functioning of my introspective mechanisms, such as bias, or trauma). For example, I might believe that I am in pain when in fact I am just extremely cold. Or in some contexts, I might mistake the feeling of extreme heat with the feeling of extreme cold. But as Bettcher explains, drawing again on a consensus in philosophy of mind, these cases seem to be highly unusual, so that for practical purposes, we can assume that first-person present-tense avowals of phenomenal states are virtually incorrigible. It would be very weird, from an epistemic point of view, for someone to challenge my judgements about my own phenomenal states. It is true that I am not 100% certain, but certainty is not required for knowledge nor strong justification.

Furthermore, even if we accept that it is possible that my avowal of my own phenomenal state is false, we can still accept the following. Imagine that after my *asserting* that I have a phenomenal state, another person *asserts* that I am wrong, or doubts that my judgement is right. This seems to generate a strong epistemic duty from the latter speaker, namely, the speaker should provide some reasons for why the former subject's own phenomenal judgement is wrong. That is, the general fact that we are not 100% certain about our phenomenal judgements does not seem enough justification for someone else to *make the assertion* that my own judgement that I am in a given phenomenal state is not true. And it is difficult to think about what reasons could possibly be available to a third person but not available to the subject of the phenomenal state itself, at least in normal conditions.

Hence, I agree with Bettcher that we can treat phenomenal avowals as virtually incorrigible, at least for practical purposes, in the sense that the default option should be to trust phenomenal avowals, in normal conditions. We could add that in some exceptional circumstances, the hearer of a phenomenal avowal who then goes on to assert that the avowal is not true could possibly discharge their epistemic duties. That is, the speech act of doubting my phenomenal avowal would be an *appropriate* speech act only if the speaker is able to present very strong reasons for why the first-person phenomenal judgement might be incorrect, and *in addition*, why the hearer is in a better position to know than the subject herself. Again, as Bettcher argues, this is not inconceivable (for example, the subject might be under the effects of drugs that could be blocking their introspective mechanisms, and all of this might be known by the hearer), but this is unusual.[1] For this reason, it seems that to deny

1 It could be argued that our introspective judgements about our own phenomenal states are not as reliable as I am assuming (see Schwitzgebel, 2006). In any case, my main claim regarding phenomenal avowals is that it is often inappropriate for a third person to doubt a phenomenal avowal, even if these are not *incorrigible* strictly speaking. In addition, given that Bettcher is happy to accept the incorrigibility of phenomenal avowals, as opposed to the incorrigibility of avowals of standing mental states, I am happy to assume the former for the sake of the argument and focus on the discussion of the latter.

a phenomenal avowal is in most contexts an *inappropriate* move in a conversation, given the very nature of phenomenal states and our knowledge about it (namely, by means of directly introspecting our own conscious phenomenal states).

Let's now examine the alleged incorrigibility of our avowals of *standing* mental states. As Bettcher argues, there are no good reasons to claim that our judgements about our own standing mental states are incorrigible, not even with respect to present-tense avowals: "In attitudinal cases, FPA is less strong" (p. 100). And the main reason is that our knowledge of standing mental states is not by introspection alone. How we come to know our standing mental states is a complex subject, but a plausible position that is widely endorsed has it that knowledge of those mental states is achieved by means of inferential mechanisms, drawing on several pieces of evidence, including introspection of possible conscious manifestations of those dispositional mental states, observations of manifested behavior, past patterns of behavior, other standing mental states about which the subject could have independent reasons for having, and so on.[2] Hence, our knowledge of our standing mental states is based on inferential reasoning, not on introspection alone. Clearly, inferential reasoning on the basis of observable behavior is less reliable than introspection. Still, we can assume for our purposes here that we can have strong justification of this sort for our judgements about our own standing mental states, such as beliefs and desires. However, this falls short of *incorrigibility*, even the slightly modified sort of incorrigibility that we argued above is plausible to ascribe to phenomenal judgements (that is, the default trustworthiness of phenomenal avowals, given the primacy of introspection with respect to third-person epistemic methods concerning phenomenal experience), since this is not so plausible with regard to standing mental states. So, if our judgments about our standing mental states are not incorrigible in any interesting sense, the crucial question becomes whether we still have any interesting kind of *epistemic advantage* from the first-person point of view, compared to the third-person point of view.

Bettcher argues that there is no interesting sort of *epistemic* advantage from the first-person point of view. And the main reason (as I reconstruct her argument) is that given the way we come to know about our own *standing* mental states, the first person and the third person points of view seem pretty analogous, with regards to the forms of access.

Here, we should be careful. It has been established that our judgements about our own standing mental states are not *incorrigible* in any substantive sense, since we could be mistaken about them to the same extent as we are mistaken about many other judgements we make about the external world. But this gives rise to the crucial question of whether we have any epistemic *advantage* from the first-person point of view in order to know my own beliefs and desires, as opposed to the third person point of view, even if it does not amount to incorrigibility. It is clear that first-person judgements about my own standing mental states can be wrong, but can we say that first-person judgements have an *epistemic advantage* with respect to third-person judgments when it comes to knowing a person's attitudinal mental states?

As we said above, the inferential reasoning by means of which we come to know about our (standing) beliefs and desires is based on a combination of different pieces of evidence, including introspection of any conscious manifestations of these dispositional states that there might be. And when it comes to the introspection of our own phenomenal states, it seems clear that we do have an epistemic advantage from the first-person point of view as opposed to the third person, due to the mechanisms of introspection. As I put it above, phenomenal avowals are virtually incorrigible for this reason, since it would be inappropriate in most

2 See Gertler, 2011 for a useful overview.

contexts, for other people, to correct my avowal of a phenomenal state, unless they discharge their epistemic duty to offer sufficient reasons for making such an unusual assertion. The crucial question, then, is whether we can say something similar about avowals concerning standing mental states such as beliefs and desires.

Bettcher claims that we cannot talk about *epistemic advantage* in any substantive sense regarding standing mental states. In response, I think that we are justified in positing at least some sort of epistemic advantage from the first person point of view as compared to the third person, given the fact that the inferential reasoning that we use to come to know about standing mental states also relies in part on *introspection* (of conscious manifestations of our attitudinal mental states), and we do have some first-person epistemic advantage regarding introspection. Of course, Bettcher is right that this does not entail incorrigibility in any interesting sense: there is no sense in which we cannot be mistaken about our standing mental states, from the first-person point of view, as opposed to the third person point of view.[3]

However, I wish to argue that there is a sense in which it is still *inappropriate* for an external person to challenge or doubt my avowal of a standing mental state, unless they supply sufficient additional evidence that overrides my epistemically privileged access to some partial evidence for my avowal, namely, my introspective judgements. It could be the case that someone has evidence based on my observable behavior (about which another person could have more reliable records than myself, for instance), so that this overrides the partial introspective evidence that I have about whether I have a given belief or desire. Still, it can be argued that in order to challenge a first-person avowal of a belief or a desire, an external person has the *epistemic* duty to present sufficient overriding evidence. And this additional duty is precisely what explains the degree of epistemic advantage that the first-person point of view does have concerning our own standing mental states, in my view. That is, if someone were to doubt or challenge a *third-person* claim about someone else's attitudinal states, the former would have the standard epistemic duty to be able to present reasons for their assertion, but in the case of doubting or challenging a *first-person* claim about the subject's own attitudinal mental states, the speaker who challenges this claim would incur in an *additional* epistemic duty to offer reasons for why the doubter knows better than the subject herself, in this particular case.

However, this might not be enough to talk about first-person epistemic *authority* or *privilege* in any substantive sense. Bettcher argues that from an epistemic point of view, avowals of beliefs and desires do not display first-person *epistemic authority*, because they rely on inferential knowledge on the basis of observable behavior, and this is fallible. But still, she argues that it is *wrong* in some substantive sense to challenge or to doubt those avowals. Why? According to Bettcher, the reasons are *ethical* rather than epistemic. And it does not have to do so much with the kind of *evidence* the first person has for their own mental states (which is similar to the kind of evidence available to the third-person), but rather with the kind of *speech act* that we make when we self-report our own standing mental states such as beliefs and desires and other propositional attitudes. Bettcher's crucial claim is that a self-report of a standing mental state is not merely a report of a *fact* (p. 102), for which both the first person and the third person could in principle have similar sorts of evidence, but rather constitutes a different kind of illocutionary speech act, other than reporting a fact.

3 Logue (2021) also argues that our beliefs about our gender identities are not incorrigible, since they can be mistaken. She considers the example of a trans woman who did not believe she was a woman at a certain time before transitioning. Logue argues that it is important to be able to say that she was never a man, and therefore, her belief before transitioning that she was a man was mistaken (2021, p. 130).

This speech act is different in several inter-connected ways. First, when one self-reports a mental attitude such as a belief or a desire, one is *taking responsibility* for those attitudes. Bettcher argues that propositional attitudes can be appraised or criticized, for being true or false, rational or irrational, appropriate or inappropriate. And when a subject reports her own mental attitudes, she is not just describing a certain fact of the matter, given her evidence, but she is also putting her own attitudes "out there" to be publicly scrutinized, so that she can be subject to praise or criticism because of their rationality or appropriateness. Because of this, there is an important asymmetry between first-person and third-person attitude reports. And Bettcher argues that it is wrong for someone to assume that they are allowed to put someone else's attitudes out there for public scrutiny. This is something that (at least in normal circumstances) only the owner of the attitudes is entitled to do. As Bettcher explains, our mental attitudes are *private* in the sense that it is wrong for others to make them public without our consent. For this reason, to claim that one has expertise about another person's mental attitudes is ethically problematic, precisely because we are making something public that is "none of our business", that is, only the first person is entitled to make them public. Hence, this explains the special illocutionary force of attitudinal self-reports: we are making our attitudes public, attitudes that other people should not make public without our consent, and furthermore, we are putting our own attitudes out there for public scrutiny, to be appraised or criticized, in a way that makes us *responsible* for their merits or flaws. This is the sense in which the first person is an *authority* about their own mental attitudes: not because she has *epistemic* expertise over them, but because she is *authorized* (unlike external subjects) to make them public and to take responsibility for them in making them subject to rational scrutiny by others. In other words: we are entitled to becoming responsible for our own attitudes, but other people are not entitled to make us responsible for our own attitudes.[4]

Bettcher (2009) advocates for a change of the meaning of the terms 'man' and 'woman' (in particular, for a shift in our patterns of usage of such terms and cultural practices involving these terms). According to Bettcher, the way of using these gender terms in mainstream linguistic communities is closely related to the mainstream cultural practices about gender presentation. In the same way that gender presentation (e.g. clothes, hairstyle, etc.) is supposed to represent genital status, the ascription of gender terms such as 'man' and 'woman' are also supposed to represent people's genital status. But Bettcher argues that genital status should remain *private*, that is, it is wrong for others to expect us to reveal something that is not really "their business" to know. For these reasons, we should change both norms of appearance surrounding gender presentation and norms of use regarding gender terms. We should change the meaning (and use) of gender terms so that they do not aim to reveal genital status but rather are used to communicate someone's gender *self-identification*. This clearly respects the First-Person Authority principle about gender, according to Bettcher. Let me explain why: If gender terms aim to refer to someone's self-identification, then when we talk about someone's gender, we talk about how they self-identify. And it is clear that claims about how one self-identifies display first-person authority in the ethical sense: it would be wrong for someone else to claim expertise about our own gender self-identification. This is not due to the fact that the subject could never be wrong

4 This claim by Bettcher has intuitive appeal, although on reflection, there seem to be some problems lurking. For example, we might be entitled to make other people responsible for their propositional attitudes, such as sexist or racist or homophobic propositional attitudes. Furthermore, there might be exceptions to the restrictions to make my attitudes public: for example, my therapist might be *authorized* to make some of my hidden attitudes public (to me). Or a very close friend of mine might be authorized to tell me about some of my hidden attitudes, such as my homosexual sexual desires which I want to repress due to internalized homophobia. Thus, there are exceptions where other people might also be allowed to make someone else's attitudes public and subject to scrutiny.

about her own self-identification (because she could be wrong), but rather because her *speech act* of declaring her gender is thereby a speech act of declaring her gender self-identification (assuming gender terms come to refer to gender self-identification), that is, a gender avowal on this understanding becomes an act of the subject making public, and taking responsibility for, the mental attitude of self-identifying with a gender. And as Bettcher argues, other people are not *allowed* to disclose my own mental attitudes without my consent. This is what the ethical reading of first-person authority involves.

Now, this argument still gives rise to an epistemic worry, as follows. Bettcher has clearly established that it is ethically problematic for another person to *claim* that they have expertise about my gender self-identification, because this denies my autonomy, in the sense that it violates an ethical principle according to which other people are not allowed to reveal something that is not up to them to reveal. Nonetheless, could it be argued that someone else could have a *belief* about my own gender self-identification that is epistemically better justified than my own belief about it? Bettcher's argument concerns *assertions* about someone's gender, and she argues that the *speech act* of gender avowal is different from the speech act of revealing somebody else's gender, so that only the former speaker, but not the latter, is *authorized* to perform such a speech act.[5] This ethical argument concerns *speech acts*.

But what about the justification of our *beliefs*, in cases these beliefs are not linguistically expressed? The central question that arises is whether my first-person *beliefs* about my attitudinal mental states have any *epistemic advantage* over third-person beliefs about someone else's mental states. It is not clear to me what Bettcher's argument would entail about this issue. Since this is an important question, more attention should be paid to this issue.[6] In the next section I will explore an alternative account that offers an answer to this question, an answer that I will argue is misguided. And in section III I will develop what I take to be a more satisfactory answer.

II. Fileva's account

Iskra Fileva (2021) has recently offered a novel account of gender, according to which the features that constitute gender are all knowable by introspection. If this account were correct, we would have an answer to the question I posed in the previous section, to wit: if gender can be known by introspection, then this explains why my *belief* about my own gender is *epistemically privileged* as compared to other people's beliefs about my own gender. As we saw, on Bettcher's own account, gender is based on attitudes of self-identification. And since, according to Bettcher, my beliefs about my own attitudinal mental states are not epistemically privileged, as compared to other people's beliefs about my own attitudes (for we all come to know about my own attitudes in the same way, that is, by inferential reasoning), then it follows that my beliefs about my gender are not epistemically privileged. This is why Bettcher defends the FPA principle as an *ethical* claim, not as an *epistemic* principle. But Fileva disagrees: she claims that our gender is *constituted* by aspects of our experiences that are knowable by

5 Does this mean that no-one else is authorized to use gender pronouns and gender terms to reveal my gender identity? This could be a possible consequence of Bettcher's ethical reading of the FPA, but this is not how I am understanding it. Bettcher's view, at its core, is that a first-person gender avowal is a speech act of making the subject's own attitudes public, whereas a third-person gender ascription is merely the speech act of describing a fact, which can be true or false, justified or unjustified, and so on. This is the asymetry between the first person and the third person: it is not epistemic, but rather ethical, having to do with the evaluation of what the speaker is doing, from a normative point of view (i.e. describing someone's attitudes vs. making my own attitudes public).

6 Turyn (2023) argues that we have ethical reasons to advocate for an epistemic version of FPA. One main reason is that since it is possible to wrong someone because of what we believe about them, it can be argued that to fail to believe someone's gender avowal can constitute a wrong, and therefore, we ought to believe what others say about their gender identities.

introspection. Her account is a two-tier account. First, she poses a necessary and sufficient condition for someone to have a gender: S has gender G iff S would identify with gender G in ideal circumstances (for example, free from oppression) (2021, pp. 189-190). But she acknowledges that circumstances are not always ideal. Then, she wonders what features of our experiences would give rise to our identifications with a certain gender in ideal circumstances, and she claims that the facts of experience that give rise to gender identification include the following: (a) "comfort or discomfort with the sexual characteristics of one's body" (p. 190); (b) identification with and affinity for people of the sex you identify with (for example, ciswomen would identify with people of the same sex, transwomen would identify with people of a different sex than the sex they were assigned at birth, etc.); (c) " a substantial number of ... preferences about clothing and appearance" (p. 190); and (d) "a strong and persisting preference for being recognized and treated as a man or a woman" (p. 190).

Fileva realizes that none of these criteria is *individually necessary* since there can be people who identify as, say, women, but do not satisfy some of these criteria. For example, many women do not have a preference for clothing and appearance associated with being a woman but can actually have a strong preference for clothing and appearance associated with being a man (as in the case of butch lesbians, or gender non-conforming women). Also, many women do not have a strong affinity with other women but rather prefer the company of men. Fileva responds to this worry by claiming that criteria (a-d) should be understood as a *family-resemblance concept* rather than in terms of necessary and sufficient conditions. In particular, Fileva claims that none of the criteria (a-d) is individually necessary, and furthermore, she adds, "the question is left open of whether they are jointly sufficient (my account allows for the possibility that they are not, though I consider this possibility highly unlikely)" (2021, p. 194). Also, since the account should be understood in "family-resemblance" terms, one could ask how many criteria are necessary in order to satisfy the account. Fileva answers that this is an empirical question. We can conclude, then, that on this account, satisfying an unspecified number of the criteria (a-d) is necessary, where what counts as a sufficient number of criteria is left open, and where no single criterion is individually necessary. Moreover, satisfying all the criteria may not be jointly sufficient to satisfy the account, although this possibility is deemed highly unlikely.

Unfortunately, I think this account faces some serious worries. My main worry is that there could be women who satisfy none of these criteria (a-d), and they still count as women, so it follows that satisfying a sufficient (even if vague) number of criteria is not necessary. Fileva's central claim is that someone is a woman if and only if, in ideal circumstances, she would identify as a woman. Given that this claim by itself does not make the epistemology of gender any easier (since we are often not in ideal circumstances), she aims to offer additional criteria that would make the epistemology of gender easier, and this is why she offers criteria (a-d). Also, she seeks to address the question of the basis on which we identify with a particular gender such as being a man or a woman. To this, she answers: on the basis of (a sufficient number of) criteria (a-d). As we have seen, Fileva also acknowledges that the criteria (a-d), taken together, are not *jointly sufficient*, that is, someone can satisfy *all* these criteria related to being a woman, say, because of societal pressures, without being a woman (since she would not identify as a woman in ideal circumstances), although this is highly unlikely on her view. Furthermore, she argues that satisfying a sufficient (even if vague) number of these criteria is *necessary* for being a woman, even if it is not sufficient for being a woman.

I disagree: it seems to me that the *disjunction* of criteria (a-d) is not necessary either, that is, someone can be, say, a woman, even if she does not satisfy *any* of these criteria (but would still identify as a woman in ideal circumstances). Given that the criteria are neither jointly sufficient nor disjunctively necessary, it is not clear that this really counts as a family-resemblance account. Rather, criteria (a-d) could be understood as a set of *fallible guides* for

knowing someone's gender: if someone has the experiences related to those criteria, it is likely that they will be of the corresponding gender, but this guide might be fallible in many cases. Fileva agrees that the criteria (a-d) might be fallible, but she also claims that criteria (a-d) provide a family-resemblance account of gender, which as I have argued does not seem correct. Furthermore, she claims that criteria (a-d) "are introspectively accessible aspects of experience that are constitutive of gender and simultaneously underwrite our knowledge of our own gender" (2021: 190). And she adds: "dysphoria or the absence thereof does not simply provide evidence of gender, it is partly *constitutive* of it" (190). In response, it seems controversial that criteria (a-d) are *constitutive* of gender, since those criteria are not jointly sufficient for having a certain gender, and furthermore, someone could have a certain gender without satisfying any of the corresponding criteria, that is, the criteria (a-d) are neither jointly nor disjunctively necessary to have a gender.

For all these reasons, Fileva's account has not adequately explained the epistemology of gender yet. If criteria (a-d) are no more than *fallible* epistemic criteria for ascertaining someone's gender (including my own gender), then what is gender-identification really based on? We can agree that criteria (a-d) are easy to know from the first-person point of view, but criteria (a-d) do not *determine* someone's gender.[7] So, we can still ask: what determines someone's gender, and is *that* knowable by introspection? On Fileva's account, the only criterion that is truly necessary and sufficient for gender is self-identification *in ideal circumstances*. But this is clearly not knowable by introspection either. So Fileva's account does not provide an answer to our question, namely, whether our *beliefs* about our own gender identities are epistemically privileged as opposed to third-person judgements.

III. A possible solution

In my view, Bettcher's account already provides the resources for an adequate explanation of the epistemology of gender self-identifications, including how our first-person *beliefs* about gender identity (in addition to our first-person gender *avowals*) can be considered privileged. The reason was already suggested in sect. I: our knowledge of our attitudinal mental states is based in part on introspection of the *conscious manifestations* of these very dispositional mental states. For it is part of the nature of propositional attitudes such as beliefs and desires that they can be *manifested*, under certain manifesting conditions, in the form of conscious mental states. Beliefs and desires also have other *non-conscious* manifestations such as causing behavior and other attitudinal mental states. But, given the strong connection between a disposition and its manifestations, the fact that we can know some of the dispositions' manifestations by *introspection* explains why I can have an epistemic advantage over other people who are not able to know the conscious manifestations of *my* mental attitudes by introspection.

The central question above was whether we can say that our *beliefs* about our own gender are, in any meaningful sense, *epistemically privileged*—even if we do not express them linguistically, and therefore, they are not subject to Bettcher's ethical FPA principle. My answer to this question is that our standing mental states of identifying with a certain

7 Fileva claims that criteria (a-d) can be known by introspection, but this is not clear. Most of the criteria appeal to *attitudinal* mental states, such as identification and affinity (b), preferences about clothing and appearance (c), and a persisting preference for being recognized in certain ways (d). These all seem to correspond to dispositional mental states rathen than ocurring mental states. Perhaps criterion (a), about comfort or discomfort, may correspond to phenomenal mental states, but it also has a dispositional dimension in my view. Given the view developed in the previous section, these attitudinal mental states can be known by inference on the basis of multiple factors, including (but not only) the *manifestation* of the dispositions in the form of *conscious* mental states, where these conscious mental states are clearly knowable by introspection. In this way, we can be said to have a certain degree of *epistemic priviledge* with respect to criteria (a-d) from the first person point of view, even if these are not completely knowable by introspection alone.

gender can be known only by means of inferential reasoning (which in principle everyone is in a similar position to achieve). Crucially, however, this inferential reasoning is based on a combination of pieces of evidence, including the *introspection* of conscious manifestations of those very attitudinal mental states, as well as the observation of behaviors that are also a manifestation of those mental states.[8]

To conclude, we can agree with Bettcher that we are not incorrigible about our own gender judgements.[9] But, in addition to an *ethical* version of the FPA principle about gender avowals, we can also hold a modest version of an *epistemic* FPA principle, according to which we are not incorrigible, to wit: I could believe that I have gender G, when this is false, but I am typically in a much better position to know my gender identity than other people are with respect to my own gender identity. Thus, a first-person judgement about gender identity is epistemically privileged with respect to a third-person judgement about gender identity. This means that if I were to linguistically express my gender judgement and someone were to challenge it or doubt it, this challenge would come with a strong epistemic duty, that is, not only the usual epistemic duty to be able to provide evidence to justify their judgement (as with most assertions), but also *additional* evidence in order to justify why they have good reasons to believe that their third-person point of view in this particular case *overrides* my first-person point of view. I think this result nicely complements Bettcher's defense of an ethical FPA principle and can be considered as an *epistemic* version of the FPA principle.[10]

REFERENCES

Bettcher, T. M. (2009). Trans identities and first-person authority. In L. J. Shrage (Ed.), *You've changed: Sex reassignment and personal identity* (pp. 98-120), Oxford: Oxford University Press;

Fileva, I. (2021), The gender puzzles. *European Journal of Philosophy*, 29(1), 182-198. DOI: 10.1111/ejop.12550;

Gertler, B. (2011). *Self-Knowledge*. Milton Park: Routledge;

Jenkins, K. (2016). Amelioration and Inclusion: Gender Identity and the Concept of *Woman. Ethics,* 126(2), 394-421;

Jenkins, K. (2018). Toward an Account of Gender Identity. *Ergo*, 5(27), 713-744;

Keeling, S. (2021). Knowing our Reasons: Distinctive Self-Knowledge of Why We Hold Our Attitudes and Perform Actions. *Philosophy and Phenomenological Research*, 102, 318-341. DOI: 10.1111/phpr.12655;

Keeling, S. (2023). Accounting for doxastic agency: Mental action and self-awareness. *Synthese* 201, 185. DOI: 10.1007/s11229-023-04069-7;

Logue, H. (2021). Gender Fictionalism. *Ergo* 8(28), 125-162;

Schwitzgebel, E. (2006). The unreliability of naive introspection. *Philosophical Review*, 117 (2), pp. 245-273;

Turyn, G. (2023). Gender and first-person authority. *Synthese*, 201, 122. DOI: 10.1007/s11229-023-04125-2.

8 Moreover, it could be argued that, from the first-person point of view, we have better epistemic access to our own behavior—that is, knowledge of our *actions*—than to the behavior of others, about which we lack a *sense of agency*, a sense arguably accessible through introspection. (See Keeling 2021, 2023.)

9 Jenkins (2018) also agrees: whereas she claimed in her (2016) paper that we do have the genders that we believe we have (i.e. what I called *incorrigibility* above), Jenkins (2018) claims, drawing on Bettcher (2009), that not everyone who believes that they have a certain gender is right about it. Jenkins (2018) endorses the ethical version of FPA, not the epistemic version. In this paper, I aim to clarify to what extent we can defend an epistemic version too.

10 This work has been funded by grant PID2021-124100NB-I00, funded by MCIN/AEI/10.13039/501100011033 (Spanish Ministry of Science and Innovation). I am also grateful to audiences at Glasgow, Lisbon, Sheffield, Wollongong, and the Pacific APA, as well as an anonymous referee and the editors of this volume.