Gender Fictionalism

HEATHER LOGUE
University of Leeds

This paper develops a proposal about the metaphysics of gender by focusing on the question, what is it to be a woman? In recent years, the view that it is a matter of self-identifying as a woman has become increasingly popular outside of philosophical circles. Metaphysicians of gender generally regard this kind of view as hopeless, but it is the only kind of view that accommodates the strongest form of first-person authority (FPA) over gender.

This inquiry into the nature of gender is an ameliorative one, which takes the aim of securing the strongest form of FPA as its starting point. The main goal of this paper is to show that a self-identification account of gender can be made philosophically respectable, despite conventional wisdom to the contrary—if we embrace fictionalism about gender discourse.

In Section 1, I will outline the belief condition (a specific version of a self-identification account of gender), and detail several seemingly insurmountable objections to it. In Section 2, I will motivate the search for an account that accommodates the strongest form of FPA. In Section 3, I will outline fictionalism about gender discourse and explain how it can address the objections to the belief condition. In Section 4, I will flesh out some key details of gender fictionalism. In Section 5, I will outline and respond to a family of serious objections, to the effect that gender fictionalism trivializes gender. Section 6 briefly considers the question of whether we should do away with the "gender fiction".

WHAT is it to be a woman?[1] A common reply to this question is that it is simply a matter of being "biologically female", where this is supposed to be a matter of (e.g.) having female reproductive features (ovaries, uterus,

---

1. I'll focus mainly on accounts of womanhood in this paper, but analogous points will apply to accounts of other genders. Note that the view I will defend is a view about *gender*; it is neutral on whether we can or should distinguish between biological sex and gender (conceived as something along the lines of the social and/or psychological significance of sex).

Contact: Heather Logue <h.a.logue@leeds.ac.uk>

vagina, breasts), or two X chromosomes, or something else along these lines. But this reply seems wrong—there are women who fail to meet most or all of these conditions (some trans women, some women with androgen insensitivity syndrome).[2]

Philosophers who reject this common reply have offered quite a few alternatives (see, e.g., Stoljar 1995; Hale 1996; Haslanger 2000; Alcoff 2006; Bettcher 2009; 2012; Witt 2011; Ásta 2011; 2018; Bach 2012; McKitrick 2015; Díaz-León 2016; Jenkins 2016; 2018). In addition, there is an alternative that has been growing in popularity outside academic philosophy in recent years: roughly, that being a woman is a matter of *self-identifying* as one.

Although this kind of account is not widely accepted in academic circles, a certain strand of philosophical inquiry concerning gender leads naturally to it. According to Sally Haslanger, some philosophical inquiry is *ameliorative* in nature. Such an inquiry starts by asking what the purpose of a given concept is—for example, the concept of knowledge, or the concept of womanhood—and offers an account of the target phenomenon that serves that purpose (Haslanger 2005). Ameliorative inquiry concerns what our concepts *should* be like and what they *should* pick out (given our purposes), rather than what they are currently like and what they do in fact pick out. Hence, ameliorative inquiries require normative input (Haslanger 2005: 20).

One kind of normative input to ameliorative theorising about gender is the aim of accommodating some form of *first-person authority* (FPA) about gender (Bettcher 2009)—broadly speaking, the claim that a person is the ultimate authority on what their own gender is. Talia Bettcher proposes FPA as a principle that a trans-friendly ameliorative conception of gender ought to respect (2009), and Katharine Jenkins (2018: 719–20) follows Bettcher in regarding FPA as a desideratum on an ameliorative account of gender identity. Accommodating FPA about gender serves the further purpose of avoiding many harms caused by denying it. For example, Talia Bettcher describes how denying FPA can demolish a trans person's sense of self (2009: 114–15), and Stephanie Kapusta (2016: 504–5) details several psychological, moral, and political harms that result from denying FPA through misgendering trans people. Ameliorative inquiry with these normative inputs leads naturally to something along the lines of a self-identification account of gender. If self-identifying as a member of a given gender is sufficient for being a member of that gender, a person's sense of their own gender determines what their gender is. On this view, a person's authority over their own gender is absolute, and FPA over gender is thereby straightforwardly accommodated.

---

2. For a recent defence of the common reply see Byrne (2020), and for criticism of this defence, see Dembroff (2021).

However, even metaphysicians of gender sympathetic to such an account stop short of endorsing it, because they regard it as philosophically hopeless. According to Bettcher,

> If believing one is a woman . . . [is the] sole determinant of membership, there are difficulties concerning an account of what it is to believe one is a woman. Is it to believe one possesses the special feature making one a woman? If so, to believe one is a woman is to believe one believes one is a woman. And now we seem to have some problem of circularity or regress. In practice this means that the criterion is virtually unintelligible. (2009: 109)

And Jenkins argues that combining a self-identification account of gender identity with the view that gender identity determines one's gender class results in vicious circularity:

> someone who asks what it means to say that a certain person 'has a female gender identity' will be told that it means that that person has a sense of herself 'as a woman'—but if the questioner then asks what a 'woman' is, they will be told that a woman is 'a person with a female gender identity'. Thus, the questioner is none the wiser as to what it means to have a female gender identity. (2018: 714)

In short, such an account seems to be hopelessly circular. (And as I'll explain below, there are even more apparently damning objections.) For this reason, even those who hold that we ought to exhibit the utmost respect for a person's sense of their own gender in our theorising think that we cannot take the most straightforward route to that end, and as a result end up defending significantly weakened versions of FPA over gender.

In this paper, I will not add to the case for the claim that an ameliorative inquiry into gender should take as normative input the aim of accommodating FPA over gender. For the purposes of this paper, I am going to assume for the sake of argument that we should accommodate it. My goal in this paper is to argue that we should accommodate FPA in its strongest form, and that there is a philosophically sound way of doing so (despite the apparent consensus to the contrary). We can accommodate the strongest form of FPA by adopting something along the lines of the self-identification account, and we can overcome the apparently damning objections to accounts of this sort by adopting fictionalism about gender discourse.

In the first section, I will outline the belief condition (a specific version of a self-identification account of gender), and detail several seemingly

insurmountable objections to it. In the second section, I will motivate the search for an account that accommodates the strongest form of FPA. In the third section, I will outline fictionalism about gender discourse and explain how it can address the objections to the belief condition. In the fourth section, I will flesh out some key details of gender fictionalism. In the fifth section, I will outline and respond to a serious objection to gender fictionalism. The final section briefly considers the question of whether we should do away with the "gender fiction".

1. The Belief Condition

The claim I will defend is that *believing that one is a woman is sufficient for being one.*[3] Let us call this the *belief condition* on womanhood. Note that such a belief may well be dispositional rather than occurrent; it need not be a conscious, explicit judgement. On this kind of account, a subject's beliefs about their gender are self-verifying (the very existence of a belief about one's own gender makes it true), and subjects are thereby infallible and incorrigible with respect to their own gender. (One is infallible about something if they can't be wrong about it, and one is incorrigible about something if no one can be justified in believing that they're wrong about it.)

The belief condition is in the vicinity of the self-identification account sketched above. In some cases, talk of self-identification might just be another way of talking about what one believes about oneself. But sometimes talk of self-identification seems to pick out a broader phenomenon (see, e.g., Bettcher 2009: 109). I have a firmer grip on the notion of belief than I do on the notion of self-identification, which is why I am using the belief condition as a case study. That being said, most (if not all) of the problems I am about to discuss also arise for a condition framed in terms of a broader notion of self-identification.

The belief condition has the virtue of accommodating FPA about gender, and thereby avoiding the harms caused by denying it. Although this is a strong consideration in its favour, it seems to be outweighed by other, seemingly insurmountable problems.

One kind of problem that has been raised for self-identification accounts is a problem for the belief condition too. The fundamental issue is that such accounts

---

3. For simplicity's sake, I will not always carefully distinguish between gender categories and gender identities in what follows (compare Jenkins's related distinction between gender-as-class and gender-as-identity in her 2018: 406–8). This is because the gap between gender category and gender identity on the kind of account I'm defending is not very wide: to have a woman's gender identity is just to believe oneself to be a woman, and if one believes oneself to be a woman, one is a woman.

do not explain why gender is so important to us. As Katharine Jenkins puts the point,

> on this account gender identity is equated with a disposition to make certain kinds of assertions [e.g., 'I am a woman']. This means that the account makes gender identity seem trivial: why should we care about dispositions to utter certain sentences? Insofar as we care about gender identity we seem intuitively [to] care about it as *whatever it is that makes people want* to utter those sentences, or *whatever it is that they express* when they do utter them. It is not simply that the self-identification account needs to say more in order to explain how gender identity is important and deserves respect, but rather that it is difficult to see how the account *can ever* say more on this, due to its minimalist stance. (2018: 728, emphasis in original)

A similar objection could be raised for the belief condition. Insofar as we care about gender, we seem to care about it as whatever it is that prompts people to form beliefs about their genders. One's belief that one is a woman is important and deserving of respect only because of the importance of what gives rise to this belief in the first place (whatever that is, exactly).

Jenkins also objects to self-identification accounts on the grounds that they cannot explain why some trans people desire transition-related healthcare, such as hormone therapy and gender-confirmation surgery: "it is difficult to perceive any relationship at all between a linguistic disposition and the sort of felt need for one's body to be different that would prompt the desire to access transition-related healthcare" (2018: 728). Similarly, it is not clear why the mere belief that one is a woman would prompt the desire to access such treatments. Presumably, the explanation is to be found at the level of psychological facts that give rise to this belief, and so this is where we should be looking when constructing an account of gender identity.

Let us work our way into the next set of problems by beginning on a positive note, by considering a kind of problem for a broader self-identification condition that *isn't* a problem for the belief condition. This kind of problem concerns cases in which a person identifies as a given gender for a shallow, trivial reason—for example, a man who self-identifies as a woman with the sole purpose of getting discounted drinks on ladies' night at a nightclub.[4] Such a jerk is definitely not a woman, but a simple version of the self-identification account would predict that he is. We could modify the account to avoid this result (e.g., by adding a condition that the self-identification must be issued in good faith). But note

---

4. Thanks to an anonymous referee for this case.

that this issue doesn't arise for the belief condition at all—given that the guy in search of cheap drinks doesn't genuinely believe that he is a woman, the belief condition doesn't deliver the result that he is one.

That being said, it seems possible that a person's grip on the notion of womanhood is tenuous enough that they genuinely but falsely believe that they are a woman, just as Burge's patient with a pain in his thigh falsely believes that he has arthritis (1979: 77).[5] And this is just the tip of the iceberg; plausible counterexamples to the belief condition abound. For example, what if a child believes that she is a woman? My inclination is to say that this belief is false, simply because women must be adults. There are also numerous counterexamples involving inferences from dubious claims. To describe just a couple:

> (The meek-and-mild theory of womanhood) Suppose that a woman believes that being meek and mild is at least sufficient for being a woman, and on this basis believes that she is a woman. Although her belief that she is a woman happens to be true, it doesn't seem to be self-verifying, infallible, or incorrigible.[6]

> (Delusion) Suppose that I suffer from the delusion that I am David Hume, and on that basis alone, I believe that I am a man. And suppose that if it weren't for this delusion, I would believe that I was a woman instead. One might insist that this is a case in which I have a false belief about my gender, contrary to what the belief condition predicts.[7]

The general form of these counterexamples is that the subject's belief about their gender is based on a false claim: in the first case, a false theory of womanhood, and in the second case, a delusion concerning who one is. But a belief about one's gender that is based on a false claim is not self-verifying, infallible, or incorrigible.

Yet another sort of counterexample involves gender transitions. Consider pre-transition trans women who don't yet believe that they are women. Presumably, many of these women will believe that they are men. Respecting FPA over gender arguably requires that a trans woman can truly say that she *never was* a man, even though she may have falsely believed that she was once upon a time. But the belief condition doesn't allow for this—when she believed at time t that she was a man, the belief condition entails that this belief was true at t.

Another worry about the belief condition is that it leads to what some would regard as an implausible proliferation of genders. This is because by ruling out

---

5. Thanks to an anonymous referee for this case.
6. Thanks to Ali Boyle for this case.
7. Thanks to Lea Salje for this case.

mistakes about one's own gender, it also rules out mistakes about what the genders are. For example, if I believe that my gender is pizza, then my gender is pizza, and therefore pizza is one of the genders (Reilly-Cooper 2016).

Yet another objection to the belief condition is that it just seems implausible that sincerely believing that one is an X is sufficient for being an X for other instances of X. For example, if I believed that I am an Asian person, or a cat, most people would insist that I am just plain *wrong*. Now, some are open to the possibility that I'm right (see Tuvel 2017 with respect to race, and Roberts 2015 with respect to species). But these are minority views; people generally seem to be much more open to the belief condition with respect to gender than other identity categories. It's not immediately clear that this is a consistent position, though; if we think that believing that one is a cat is not sufficient for being one, why should we think that believing that one is a woman is sufficient for being one?

Finally, the most fundamental issue with the belief condition—mentioned at the outset—is that it is circular, and so hasn't gotten us any closer to understanding what womanhood *is*. The condition for being a woman makes reference to womanhood (believing that one is a *woman*). But to say that believing that one is a woman is sufficient for being a woman raises the question: what, exactly, do you believe that you are? Presumably there must be some deeper fact in virtue of which one is a woman, and which is (or at least could be) the ground of a person's belief that she is one (Bettcher 2009: 109). In short, the problem is that the belief condition doesn't seem to be an acceptable stopping point in the search for an account of womanhood.

Before moving on, I will briefly address a consideration that one might wrongly take to be an objection to the belief condition. The worry is that if claiming to be a woman is sufficient for counting as one in the eyes of the law, then men with nefarious intentions would be able to claim to be women in order to access spaces for women in vulnerable positions (e.g., rape crisis shelters). This worry has recently been pressed by Kathleen Stock (2018), specifically with reference to proposed changes to the UK Gender Recognition Act. However, this does not constitute an objection to the belief condition, because the proposition that believing that one is a woman is sufficient for being one is a metaphysical claim about womanhood, which is distinct from the practical proposition that claiming that one is a woman should be sufficient for entry into women-only spaces. We can accept the metaphysical proposition without accepting the practical one; the former doesn't entail the latter simply because it is possible to *lie* about whether one believes oneself to be a woman. For this reason, it would be extremely imprudent to say that claiming to believe oneself to be a woman should be the *only* condition required for access to a space for vulnerable women. That being said, I don't know what should be required for access into

such women-only spaces—that is an extremely difficult question. My point here is simply that the metaphysical question of which people are women is separate from the thorny epistemic question of how we identify who these people are for practical purposes.

2. First-Person Authority over Gender

As we have seen, the belief condition faces many difficult obstacles, and so it is unsurprising that metaphysicians of gender have given up on it. And one might think that the belief condition is overkill, anyway; although it seems to be the only way to accommodate the claim that one's beliefs about one's own gender are self-verifying and infallible, perhaps this claim is not required to secure FPA in the relevant sense.

Talia Bettcher offers a qualified version of the self-identification account that aims to secure a weaker form of FPA (2009).[8] First, she claims that we have *ethical* FPA over gender, as opposed to epistemic FPA. Epistemic FPA over gender is a matter of having a substantial epistemic advantage concerning what one's gender is. The strongest form of epistemic FPA amounts to a person being infallible about their gender. But it's hard to see how that could be without beliefs about one's own gender being self-verifying, and it's hard to see how *that* could be unless the belief condition is true. Instead, Bettcher argues that we should aim for *ethical* FPA over gender (2009: 101–3). On this view, a person is incorrigible about their own gender not because they are epistemically privileged with respect to that matter, but rather because questioning their gender self-ascription would amount to a violation of their autonomy—akin to the violation that occurs when a man insists that a woman wants to have sex with him despite her claim to the contrary (Bettcher 2009: 113–15).

Second, Bettcher argues that one has FPA over gender only when it is construed as an aspect of one's *existential* self-identity, as opposed to one's *metaphysical* self-identity (2009: 110–12). By 'metaphysical self-identity', Bettcher means beliefs that locate oneself in one's account of the world, including an account of what it is to be a woman. Being a woman is part of one's metaphysical self-identity just in case one satisfies one's account of what a woman is. But the claim that I am a woman, according to a particular metaphysics of womanhood, just doesn't seem to be the sort of thing that a person has FPA over (2009: 111). By contrast, existential self-identity is "an answer to the question 'Who am I?' where this

---

8. Bettcher approvingly mentions a self-identification account of gender identity in comparison with Jenkins's (2016: 396) account. I assume that the account she intends to endorse here is the one she offers in her 2009 paper, as she doesn't attempt to solve the problems she raises in that paper for an unqualified self-identification account that secures FPA in the strongest sense.

question is taken in a deep sense. . . . The question, when taken in full philosoph-ical significance means: What am I about? What moves me? What do I stand for? What do I care about the most?" (Bettcher 2009: 110) Being a woman is part of one's existential identity insofar as it constitutes one's perspective on the world; one's values, and one's sense of self and community. Existential self-identity does seem to be precisely the sort of thing one would have FPA over.[9]

Indeed, although Bettcher doesn't explicitly say this, it seems that we have *epistemic* FPA with respect to existential self-identity. There is no gap between being a woman in the existential sense and believing that one is—the fact that one believes this just is part of one's perspective, and thereby *makes it the case* that one is a woman in the existential sense. So there's no question of getting it wrong when one's gender self-ascriptions are taken as existential rather than metaphysical claims.

To summarize: instead of insisting upon epistemic FPA over gender under-stood as an aspect of one's metaphysical self-identity, we might be able to make do with either:

(a) *ethical* FPA over gender understood as an aspect of one's *metaphysical* self-identity, or

(b) *epistemic* FPA over gender understood as an aspect of one's *existential* self-identity.

Let us consider these proposals in turn.

According to Katharine Jenkins, ethical FPA is the best we can do, and there is no reason to prefer epistemic FPA in any case (2018: 738–39).[10] I disagree on both counts. In the rest of this paper, I will argue that we can accommodate epistemic FPA (in a sense). In this section, I will argue that doing so is worth our while, as ethical FPA cannot do everything proponents of FPA over gender want it to do.

Recall that ethical FPA over gender is the claim that one's beliefs about one's own gender are incorrigible for ethical reasons: for example, because question-ing those beliefs would amount to an unacceptable violation of the subject's autonomy. According to ethical FPA, it is morally impermissible to question oth-ers' beliefs about their own genders; that is, we should not say, or perhaps even

---

9. Jenkins does not say whether she intends her notion of gender identity to be understood in terms of Bettcher's notion of existential identity.

10. Although this second claim is difficult to square with her suggestion that "an account of gender identity that entailed that everyone has the gender identity that they think they have whilst also meeting all of the desiderata [on an account of gender identity] . . . would be preferable" (2018: 739). She seems to be saying that epistemic FPA is preferable to ethical FPA, other things being equal.

think, that others are wrong about their genders (even if they *are* wrong). Notice that, in principle, *any* metaphysics of gender is compatible with this claim, and so ethical FPA doesn't actually place any constraints on the metaphysics of gender. Ethical FPA can be combined with even the most sexist or transphobic theories of womanhood. For example, a retrograde folk theory might entail that I'm not a woman because I'm not feminine enough. But one could in principle endorse this theory while also holding, on the aforementioned ethical grounds, that no one should question my belief that I am a woman. Indeed, we might accept a sexist or transphobic theory because it is an accurate description of how gender in fact functions, while insisting that our gender classification practices should not be guided by that theory because the way gender in fact functions is inherently unjust (see Dembroff 2018).

I'm not suggesting that there's no point to endorsing ethical FPA—after all, it can prevent many of the harms associated with the denial of FPA. In particular, it is well suited to protect people from harms stemming from *applications* of theories of gender. For example, a person who aims to behave in accordance with ethical FPA is less likely to inflict psychological harm on someone by misgendering them. But, for those of us in search of an ameliorative theory of gender that accommodates FPA, *ethical* FPA places no constraints on the content of that theory.

Epistemic FPA is the kind of FPA that provides such a constraint. In fact, epistemic FPA seems to have functioned as a regulative ideal for some theorists, and as a result their theories approximate epistemic FPA. For example, consider Jenkins's account of gender identity, according to which having a given gender identity is (roughly) a matter of experiencing norms pertaining to that gender as relevant to oneself (2018: 730).[11] On this theory, since the gender norms one experiences as relevant to oneself usually align with one's beliefs about one's gender, it's not often that one is wrong about one's gender identity. So even though this theory is not strictly speaking compatible with epistemic FPA (since it is possible to be wrong about one's own gender identity on this theory—see Bettcher 2016), it does come very close.

At the very least, epistemic FPA is a regulative ideal an ameliorative theory of gender may aim at but never achieve. At this point, a question arises: should we regard epistemic FPA as anything more than an unattainable regulative ideal? After all, merely approximating epistemic FPA rules out extremely sexist and transphobic theories, and ethical FPA provides protection from the harms associated with the denial of FPA. So is there any need for a metaphysics of gender that's compatible with epistemic FPA?

---

11. For simplicity's sake, this is just a rough formulation of Jenkins's account; the same point applies to the fully fleshed-out version.

First of all, given that epistemic FPA functions as a regulative ideal, it seems that we ought to accommodate it if we can, all else being equal. Otherwise, it would be mysterious why it is something ameliorative theories aim at in the first place, if actually attaining it wouldn't be the best possible result (again, other things being equal). That being said, there is a deeper question here about why epistemic FPA functions as a regulative ideal in the first place. What does it get us that merely approximating it does not?

The answer to this question is complicated, and so I cannot answer it fully here. But the basic idea is this: a state of affairs in which someone has a sincerely, deeply held but false belief about their gender is one in which that person suffers what Dembroff (2018) calls ontological oppression.[12] Ontological oppression manifests "when the social kinds (or the lack thereof) unjustly constrain (or enable) persons' behaviors, concepts, or affect due to their group membership" (Dembroff 2018: 26). Ethical FPA implores everyone to look the other way, towards a more just gender system; but it does not itself eradicate the ontological oppression that it effectively instructs us not to reinforce through our thoughts and actions. Ethical FPA is a means to the end of eradicating this kind of ontological oppression, but we cannot stop there. By contrast, the strongest form of epistemic FPA rules out scenarios in which a person has a sincerely, deeply held but false belief about their gender, and so is a natural fixed point for an ameliorative theorist aiming to articulate a more just theory of what gender should be.

The claim that such situations involve ontological oppression is not universally accepted, and arguing for it is the complicated part of the story. However, for my purposes, doing so would take us too far afield. The starting point of this paper—the normative input for its ameliorative project—is that we should accommodate FPA over gender in some form, and the claim that a situation in which someone has a sincerely, deeply held but false belief about their gender involves ontological oppression is unlikely to be denied by those who share the aim of accommodating FPA over gender.

An analogous point applies to the proposal that we can make do with epistemic FPA over gender understood as an aspect of one's *existential* self-identity. If we stop here, and say nothing about gender understood as an aspect of one's metaphysical self-identity, we've left open the possibility that our ameliorative metaphysics of womanhood excludes someone whose existential self-identity includes womanhood. This would be a possibility in which someone sincerely believes that she is a woman, but this belief is false; and again, such a situation arguably involves ontological oppression. It is a possibility that it's reasonable to

---

12. A situation of this sort (someone having a sincerely, deeply held but false belief about their gender) is described by Bettcher (2016) as an objection to the theory defended in Jenkins (2016). In a later paper, Jenkins acknowledges that her theory doesn't rule out all situations of this sort (2018: 733), but suggests that that's too much to hope for.

expect an ameliorative metaphysics of gender to rule out, and so it is worth our while to see whether we can formulate one which does that.

In summary, our ameliorative inquiry should aim for a theory that accommodates epistemic FPA over gender, understood as an aspect of one's metaphysical self-identity. But, as far as I can see, the only way to secure the result that one is infallible about one's gender is to hold that beliefs about one's gender are self-verifying—for example, that if one believes that one is a woman, then one is a woman. In other words, securing epistemic FPA requires defending the belief condition. In what follows, I'll argue that we can overcome its apparently insurmountable obstacles by embracing *fictionalism* about gender discourse.

3. Towards Gender Fictionalism

3.1. Origin of the Idea

A long time ago, in a galaxy not very far away, it occurred to me that asking about the nature of womanhood might be like asking about the nature of the Force in the *Star Wars* films. For those unfamiliar with the details of the story, the Force is "an energy field created by all living things. It surrounds us and penetrates us. It binds the galaxy together" (Obi-Wan Kenobi, Episode IV). Moreover, some people are "strong" in the Force, and upon receiving Jedi knight training, they become exceptionally intuitive fighters who can move things without touching them, control the thoughts of "weak-minded" people, and (in some cases) see the future and preserve their consciousness after death.

The nature of the Force is left relatively unspecified in the original trilogy of films (released in the late 70s and early 80s). We're told that it's an energy field, and that living things create it somehow, and we see what some people can do with it, and . . . that's pretty much it. It doesn't really bother most people that the details of the nature of the Force aren't fixed by the fiction; they're happy to just take it for granted that the Force is a thing and get on with enjoying the story. "What is the nature of the Force?" is a question that doesn't seem to *demand* an answer.[13]

Why doesn't it demand an answer? Presumably, it is because the question is about a *fictional entity*. As Kendall Walton observes, "Obviously most fictional

---

13. Although, as fans of the franchise will be painfully aware, the prequels released in the late 90s and early 00s provided one anyway. Apparently the Force consists in concentrations of subcellular microorganisms called "midichlorians", and being strong in the Force consists in a high "midichlorian count". This addition to the fiction has been widely mocked—presumably because it gives a rather ham-fisted answer to a question no one was asking in the first place.

worlds are indeterminate in many respects" (1990: 66). It is permissible to just leave the answers to some questions open; there need be no fact of the matter within the Star Wars fiction as to what the nature of the Force is. Which led me to wonder: what if we regard *gender* as a fiction? Just as we need not specify the nature of the Force within the Star Wars fiction, perhaps we need not specify the nature of womanhood within "the gender fiction". And if we need not specify the nature of womanhood with the gender fiction, the belief condition could be an acceptable stopping point after all.

It is less obvious how an appeal to fictionalism addresses the other objections to the belief condition, but I will explain how it can do so in Section 3.4. A related point is that while there may be other routes to the conclusion that we need not specify the nature of womanhood (e.g., the view that there's nothing more to genders than our practices of applying the associated terms[14]), these views don't provide responses to the *other* objections to the belief condition. Perhaps we could cobble together responses to each of the objections on a case-by-case basis, without appeal to the fictionalist framework. I'm sceptical about that, but even if we can—a matter I don't have the space to explore at length here—it would still be worth developing the gender fictionalist means to the end of salvaging the belief condition.

Note that couching the belief condition within a fictionalist framework secures epistemic FPA *only within the gender fiction*. Outside the scope of the gender fiction, one is not infallible about one's own gender. Indeed, given that gender is a merely fictional property, anyone who believes that they are a particular gender has a false belief—as there are no genders. Now, from a starting point on which securing the *non-fictional* truth of the strongest form of epistemic FPA is non-negotiable, we can argue from there to the claim that gender fictionalism is false. (Roughly: If epistemic FPA about gender is true, then my belief that I am a woman is infallible and hence true. If my belief that I am a woman is true, then genders are real properties of people, and so gender fictionalism is false.) However, this starting point is unmotivated. As I will argue in Section 5, if gender is a fiction, it is an incredibly influential one that shapes our lives in profound ways—and the value of epistemic FPA is not diminished by its realisation within *this* kind of fictional context. So there's no need to insist upon securing its truth, as opposed to its fictional truth.[15]

A few preliminary clarifications will be helpful at this point. First, talk of "the" gender fiction is (for my purposes) a harmless oversimplification. If two societies develop gender systems independently, it might be appropriate to characterize them as engaged in distinct gender fictions. And if these societies

---

14. Thanks to Simon Hewitt and Rachel Fraser for independently suggesting this view.
15. Thanks to an anonymous referee for pressing me to clarify this.

eventually come to interact with each other, at some point it might be appropriate to characterize their gender fictions as having "merged".

Second, even if the natures of genders aren't fully specified by the gender fiction, that doesn't mean that the fiction is completely silent about them. In the case of the Star Wars fiction, on any good interpretation of it, the Force is something that binds the galaxy together. This means that, according to the fiction, an essential property of the Force is binding the galaxy together. Similarly, in the case of the gender fiction, an interpretation of it will generate fictional truths about the natures of genders, even if they don't amount to a full specification of their natures. (Since it's a complex task to explain what an interpretation of the gender fiction is, and what it would be for an interpretation to be a good one, I'm not yet in a position to give an example in the case of the gender fiction that's analogous to the Force binding the galaxy together. Bear with me until Section 4.1. But for present purposes, the point is just that the gender fiction will place *some* constraints on the natures of genders.)[16]

Third, there are significant disanalogies between the gender fiction and the Star Wars example I've used to introduce it. The main disanalogy is that the Star Wars fiction is a story that audiences consume in a more-or-less passive way, whereas the gender fiction is more naturally construed as a relatively active game of "make-believe".[17] But even when someone uses the Star Wars fiction as the basis for a game of make-believe, there are significant differences between such a pretence and the gender fiction. For example, when my daughter pretended to be an Ewok last Halloween, she knew that she was pretending, the pretence was small-scale (just a few others were involved in virtue of playing along with the pretence that she was an Ewok), and it was relatively unimportant to her (it was just a way to have fun for a few hours). As I will explain in what follows, analogous claims are not true of the gender fiction.[18]

Finally, before exploring gender fictionalism further, I want to address an unfortunate coincidence. A common transphobic refrain is that transgender people are "deceivers" who are "just pretending"—for example, that a trans woman is merely pretending to be a woman with the aim of deceiving others (see Bettcher 2007 for discussion of this pernicious stereotype). The claim that gender is a pretence sounds similar to this hurtful refrain, but they are different claims. First, the notion of pretending in play doesn't entail deception. We can say that (for example) "Stephen is not trying to fool anyone, of course, when he pretends. . . . To pretend, in the sense in question, is to participate verbally

---

16. Thanks to an anonymous referee for encouraging me to spell this out.

17. Although note that, according to Walton, *appreciating* the Star Wars fiction is an activity akin to a game of make-believe (1990: 224–29).

18. Thanks to an anonymous referee for the suggestion that I emphasize these differences in terms of the Halloween example.

in a game of make-believe" (Walton 1990: 219–20). Second, on the view I am developing, *everyone* who claims a gender is best characterized as involved in a pretence.

I will proceed in the hope that characterizing *both* trans- *and* cis- gender people as involved in a non-deceptive pretence addresses the worry just raised. But I concede that, as a cisgender person, I'm not in the best position to know whether this is the case.

3.2. *Comparison with Existing Fictionalist Views*

The idea that gender is a fiction suggests a *fictionalist* approach to it. As Matti Eklund describes it, "[f]ictionalism about a region of discourse can provision-ally be characterized as the view that claims made within that discourse are not best seen as aiming at literal truth but are better regarded as a sort of 'fic-tion'" (Eklund 2011). Fictionalism is a claim about language, and the focus on language presumably stems from the plausible idea that discourse about the relevant domain is what gives rise to the fiction in the first place. But once the fiction is up and running, it is sustained by further, non-linguistic factors: for example, beliefs, and for at least some domains, non-verbal behaviours. Note that we can give a fictionalist treatment of belief that is parasitic on the treatment of assertion: beliefs in propositions expressed by assertions in a given region of discourse aren't best understood as aiming at truth, but instead as constituting a fiction.

Fictionalism is typically motivated by an anti-realist metaphysics. The anti-realist motivation for fictionalism stems from well-known worries about metaphysical commitments to metaphysically "weird" entities like numbers, possible worlds, and moral properties. These worries lead many to deny that such entities exist. But then what are we to make of our talk about them? Given that the relevant sentences really are about (or at least purport to be about) num-bers, possible worlds, or moral properties, it would seem that our claims are just false, or perhaps even meaningless.

This is where fictionalism is supposed to save the day. We can deny the exis-tence of numbers, or possible worlds, or moral properties, and yet still make sense of our mathematical, modal, and moral talk. This talk should be under-stood not as aiming at truth, but rather as constituting a useful fiction. Of course, anti-realism doesn't entail fictionalism, but fictionalism enables an anti-realist to salvage our practice of talking about such entities even though such talk is literally untrue. Obviously, unlike fictions developed for entertainment pur-poses, it seems to us that we are using mathematical, modal, and moral language in a non-fictional spirit. But according to fictionalists about these domains of

discourse, either we should change how we use this language, or we are deluded about how we are using it in the first place.[19]

It is common to characterize such fictions in terms of *pretence*. Roughly speaking, engaging in these discourses amounts to pretending that there are numbers, possible worlds, or moral properties. We engage in such pretence because it is useful, despite the fact that the utterances it involves are not literally true: for example, mathematical talk is crucial for scientific theorising, modal talk enables us to plan and strategize, and moral talk enables us to regulate our behaviour in generally beneficial ways.

Just as we can be fictionalists about mathematical talk, talk of possible worlds, talk of morality, etc., we could be fictionalists about talk of gender. On this view, to say 'Abby is a woman' would be to *pretend* that Abby has a certain property. That is, gender discourse could be understood not as aiming at truth, but rather as constituting a pretence to which the vast majority of us contribute. And just like other versions of fictionalism, gender fictionalism doesn't deny that it *seems to us* that we are trying to say something literally true when we use gender discourse. Rather, it holds that we should stop trying to do this, or that we are mistaken about what we are trying to do in the first place. (To be clear, on this view, some claims expressed in gender discourse *are* literally true: e.g., 'Abby is not a woman'. But this is literally true not because Abby is non-binary, but rather because there are no women.)

There are a couple of important disanalogies between gender fictionalism (at least as I am developing it) and the other varieties of fictionalism just sketched. First, the motivation for gender fictionalism differs from the typical motivation for existing fictionalisms. As noted above, the latter are typically motivated in large part by anti-realism about the entities in the domain at issue, which is in turn motivated by worries that such entities are too weird for a respectable ontology. By contrast, gender fictionalism is not motivated by a worry that genders are metaphysically weird. At least *prima facie*, the existence of women would be no weirder than the existence of things like tables or money, and much less weird than the existence of things like possible worlds. Rather, the motivation for a fictionalist anti-realism about gender offered here is securing the strongest form of epistemic FPA over gender by way of rescuing the belief condition. Given that a fictionalist anti-realist framework affords the best responses to the objections to the belief condition, we have good reason to accept that framework (including the claim that our gender discourse is not literally true).

Note that this type of motivation for fictionalism doesn't generalize to other social categories (e.g., race). An instance of this type of motivation for

---

19. This contrast concerns the distinction between hermeneutical and revolutionary fictionalism, which I will elaborate in Section 4.3.

fictionalism is only as compelling as the belief condition it's supposed to rescue, and (as noted above) belief conditions for membership in other social categories are implausible.

Second, it is not immediately obvious that pretending to have genders would be useful, much less anywhere near as useful as pretending that there are numbers, possible worlds, or moral properties. However, gender arguably serves the function of providing a *unifying principle* for our social identities (Witt 2011).[20] The basic idea is that gender is a fundamental social role, in that it organizes and inflects all other aspects of our social identities (e.g., wife, mother, philosopher, etc.). Sometimes this is reflected by the fact that we have gender-specific terms and concepts for the social roles in question (e.g., wife, mother). Other times the way in which gender organizes and inflects social roles is less readily apparent, but becomes clear once you think about it (Witt 2011: 93–97). For example, philosophy involves philosophers arguing with each other, and philosophers who aren't men are generally held to different norms on this score (one and the same approach can come across as incisive when taken by a man, but as bitchy when taken by a woman).

It should be acknowledged that race, class, or disability (for example) could be more fundamental to someone's social identity than their gender—particularly if such a feature is made especially salient by oppression. Also, it might be the case that we can't disentangle gender from all other aspects of one's social identity, and so the identity-unifying principles are irreducibly intersectional.[21] But even so, gender still plays a large role in unifying a lot of peoples' social identities.

The claim that gender is a unifying principle of our (largely non-fictional) social identities suggests that gender is a real feature of people. However, the gender fictionalist can endorse a claim in the vicinity: namely, that gender *discourse* affords a unifying principle for our social identities. For example, we can say that my *calling myself* a woman organizes and inflects all of my other social roles. A shorthand way putting the idea is that the function of gender discourse

---

20. I should note that I'm distilling this claim out of a rich, nuanced account, which involves a commitment to a theory of gender (Chapter 2) that I don't accept, and to an ontology of social individuals (Chapter 3) I'm inclined to reject. However, it seems to me that this claim (or at least something in the vicinity of what Witt means) is in principle separable from the other aspects of her account (contrary to what Witt seems to be suggesting on p. 31). Of course, given that her argument for the claim in Chapter 4 is predicated on these other aspects of her account, I can't appeal to that argument. But (at least as I'm understanding the terms used to express it) the idea that gender provides a unifying principle for our social identities is a plausible empirical claim, and so isn't the sort of proposition that stands in need of armchair arguments anyway.

21. Thanks to Esa Díaz-León for raising these issues, and see Witt (2011: 97–104) for a defence of the idea that gender is the fundamental social role.

is to serve as "identity glue".[22] This is what I propose we get out of gender discourse, despite much of it being untrue.

Admittedly, gender discourse isn't *essential* for this purpose. Again, we have a disanalogy with some standard forms of fictionalism. For example, it would be exceedingly difficult if not impossible to do science without talking about numbers, but in principle we could get by just fine identity-wise without talk of gender. It seems possible to have a social identity with a different kind of unifying principle (e.g., race, disability, class, or even something that's actually rather peripheral, such as one's astrological sign). Perhaps it's possible to have a social identity without a unifying principle at all.

However, fictionalism about a given discourse need not be motivated by the *impossibility* of getting by without it. Even if there could be social identities unified by something other than gender discourse (or nothing at all), it's still plausible that gender discourse performs this function contingently. So there is something that we get out of engaging in gender discourse, even though in principle we could have gotten it in some other way. In practice, we don't have direct control over how our social identities are organized anyway; the formation of social identities is regulated by longstanding, deeply-rooted cultural norms. If gender discourse in fact functions as "identity glue" for many of us, then its centrality to this function would be difficult to dislodge.

3.3. *Comparison with Views of Gender in the Vicinity*

The idea that gender is a fiction seems rather similar to some extant theories of gender. For instance, one might think that it's basically the same as the well-known claim that gender is socially constructed. However, this is not the case. The claim that gender is socially constructed can be understood as a version of gender realism, on which genders are real properties of people that emerge out of social practices (Barnes 2017). But since gender fictionalism is a version of gender *anti*-realism, it is incompatible with this way of understanding the social construction claim (and hence is distinct from it). Of course, the gender *fiction* is plausibly a real entity arising out of our social practices—but it does not follow that the genders posited within it are real, socially constructed properties of people. For this reason, a fictionalist anti-realism about gender is compatible with realism about gender *structures*, that is, the facts in virtue of which people are gendered (Barnes 2020). According to gender fictionalism, these facts are facts about the gender fiction. In short, the claim that gender is socially constructed is a very broad claim, one which leaves open *how* gender is constructed. The

---

22. Thanks to Indrek Reiland for this phrase.

view being developed here appeals to the resources of a fictionalist framework in addressing this question.

This fictionalist framework seems similar in some respects to Judith Butler's well-known account of gender (1989). In particular, Butler holds that attempting to specify the nature of womanhood is a misguided endeavour (1989: 43). This is because, on her view, gender is performative: it is not a property of a person with an identifiable essence, but rather an activity one engages in (1989: 179). The claim that gender is a kind of performance doesn't seem far off from the fictionalist claim that gender involves pretence. According to Butler, performances of gender give rise to the illusion that genders are real (1989: 42)—which seems very much like the claim that gender is a fiction. On the other hand, the motivation I've offered for gender fictionalism is in tension with some of what Butler says: the idea that simply believing that one is a woman is sufficient for being one is at odds with the idea that gender must emerge out of certain kinds of activities (e.g., dressing a certain way).[23]

Unfortunately, a thorough comparison of gender fictionalism with Butler's view would take us too far afield. The aim here is just to secure the strongest form of epistemic FPA by rescuing the belief condition, using the resources of fictionalism as developed by analytic philosophers for use in other domains. It is certainly possible that much of what I'm saying may just end up being a translation of some aspects of Butler's view into a framework more familiar to analytic philosophers. Even if that's all I'm doing, it still seems like something worth doing—cashing out the theory within a different framework could illuminate new options for developing and defending it.

3.4. *How Fictionalism Rescues the Belief Condition*

Let us now return to the matter of how fictionalism can defuse the objections to the belief condition. I've already explained how it helps with the worry that the belief condition isn't an acceptable stopping point in theorising about womanhood (because it doesn't tell us what womanhood is). In brief, fictionalism justifies a refusal to specify the nature of what a woman believes herself to be. But what about the other objections?

Let's address the objections inspired by Jenkins's criticisms of self-identification accounts first, as my responses to them actually don't require the fictionalist apparatus. The first worry was that the belief condition cannot explain why gender identity demands respect; we need an *explanation* of why a person's belief

---

23. My characterization of Butler's view is heavily reliant on Mari Mikkola's very helpful summary (2017).

about their gender is important and deserving of respect, and gender identity plausibly consists in the facts we would appeal to in giving this explanation. However, it's not clear why Jenkins insists that an account of gender identity *in and of itself* explains why it demands respect, because it's not obvious that gender identity is intrinsically important. Recall (from Section 3.2) Witt's point that gender plays the important role of unifying many peoples' social identities, and my point that it is contingent that gender plays this role (some other feature could have played this role instead, or perhaps it is possible to have a social identity without a fundamental organizing principle). We can explain why gender identity is important and demands respect simply by noting that, as a contingent matter of fact, it plays this important role. Indeed, if gender identity didn't play this role, it wouldn't be so obvious that it's important and worthy of respect. We can give a similar response to Jenkins's worry about explaining the desire for transition-related healthcare. It's not clear why Jenkins insists that an account of gender identity *in and of itself* explains why some trans people desire transition-related healthcare. After all, we could appeal to the very same gender norms as Jenkins does in explaining such desires (2018: 743–44), without claiming that these norms figure in the correct account of gender identity.

Let us now turn to the counterexamples to the belief condition. One option is to qualify the belief condition in order to avoid them. For example, in response to the child counterexample, we could say that if one is an adult, then believing that one is a woman is sufficient for being one. And in response to the counterexamples involving inferences from dubious claims, we might say that if one's belief isn't inferred from a dubious claim, then believing that one is a woman is sufficient for being one. However, it's not obvious that we can do this in response to all counterexamples. And even if we could, a suitably qualified interpretation might quickly become rather unwieldy.

Fictionalism affords a more unified response. We can simply "declare offending fictional truths deemphasized" (Walton 1990: 182). Compare what Kendall Walton says about engaging with the painting *The Last Supper*: "It is not easily denied that fictionally, in *The Last Supper*, the diners are lined up on one side of the table. But this fictional truth is an unimportant one, one that is not to be dwelt on or even noticed particularly" (Walton 1990: 182). Similarly, in the case of the gender fiction, we can accept the odd counterintuitive fictional truth generated by the belief condition, but insist that we shouldn't dwell on it. So, for example, if I believe that I am a man because I believe that I am Hume and that Hume is a man, we can just accept that it is true in the gender fiction that I am a man. But since this is a *fictional* truth, we can just ignore it, just as we ignore weird fictional facts in other fictions.

To be clear, this response should be invoked only for fictional truths that are unimportant—like my being a man because I believe that I'm a man, which

I believe *solely* because I'm under the delusion that I am Hume (rather than because, e.g., I feel like manhood is a fundamental part of who I am). This response is not appropriate for fictional truths that seem counterintuitive to some people, but are regarded as very important to others.

This point applies to the worry that the belief condition leads to an implausible proliferation of genders. Recall that, according to the belief condition, if I believe that my gender is pizza, then my gender is pizza, and therefore pizza is one of the genders. Now, if we couch the belief condition within a fictionalist framework, if I believe that my gender is pizza, then it is a fictional truth that pizza is one of the genders. If I sincerely believe that my gender is pizza, and being of this gender is deeply important to me, this is not a fictional truth that should be de-emphasized or ignored. However, I recognize that some people will regard this this fictional truth as counterintuitive. If they shouldn't ignore this fictional truth, why shouldn't they regard it as a reason to reject the belief condition even when couched in a fictionalist framework?

They shouldn't regard it as a reason to reject the belief condition because their intuitions about gender claims that don't directly affect them shouldn't dictate whether a fictional truth is acceptable or not. If I sincerely believe that my gender is pizza, and being of this gender is deeply important to me, then— given that gender is a fiction—the fact that *other* participants in the pretence find this claim counterintuitive shouldn't carry much weight. If someone insists that pizza being a gender is an unacceptable fictional truth simply because they find it counterintuitive, despite the fact that someone else derives something of great value to them from that claim, their stance is uncomfortably akin to that of a schoolyard bully who insists upon setting the terms of a game of make-believe regardless of the other children's interests. In general, when considering whether a given claim is acceptable as a fictional truth, we have to weigh its counterintuitiveness to some against its attractiveness to others, and the latter will outweigh the former if the claim is deeply important to some people.

Let's now return to the counterexample involving the pre-transition trans woman who believes that she is a man. On the one hand, during the pre-realisation period when a trans woman is claiming to be a man, accommodating epistemic FPA seems to require taking her at her word. On the other hand, post-realisation, accommodating epistemic FPA seems to require that we take her at her word when she insists that she never was a man. But it seems like we cannot do both, on pain of inconsistency; it seems that we must either say that her earlier, genuinely held belief that she is a man was false, or that her later, genuinely held belief that she never was a man is false.

One option is to ignore the inconsistency, just as we ignore other unusual fictional facts. But one might worry that this doesn't fully respect the trans woman's post-realisation insistence that she never was a man. Fortunately,

fictionalism affords a more nuanced response to the counterexample in terms of "retconning"—*ret*roactively ensuring the *con*tinuity of the fiction.[24] That is, we can stipulate that something that *was* true in the fiction at an earlier time has *always* been false. During the pre-realisation period, it was true in the fiction that S was a man (in accordance with what she genuinely believed about herself). But after S realizes that she was a woman all along, we can stipulate that it was never true that she was a man.

Note that the retconning possibility afforded by gender fictionalism is an independent consideration in its favour, given that we should accommodate the trans woman's insistence that she never was a man. For it's not clear how any other account can do that. For example, recall Jenkins's account of gender identity, according to which having a woman's gender identity is (roughly) a matter of experiencing norms pertaining to women as relevant to oneself (2018: 730).[25] If the trans woman experienced norms pertaining to *men* as relevant to herself at an earlier time t, then, according to Jenkins's account, she had a man's gender identity at t—any later claims to the contrary notwithstanding. And it seems that we'll get the same result on any account that isn't couched within the fictionalist framework: it will place conditions on manhood that the trans woman from our example is likely to meet at some point during the pre-realisation period, and conditions on womanhood that she is unlikely to meet at some point during that period. In other words, gender fictionalism is the only clear way to secure *diachronic* FPA about one's own gender.

If gender is a fiction, then retconning and allowing counterintuitive fictional truths are within our power; as Walton observes, "[p]ainters, authors, and other artists [and presumably pretenders] are veritable gods vis a vis fictional worlds" (1990: 193). Of course, as with all powers, we must be careful not to abuse them. Just as in cases involving series of books, films, and TV shows, allowing counterintuitive fictional truths and retconning should be options of last resort—when there is an important benefit that cannot be realized in any other way.[26] But it seems like the present case fits that description, given that we ought to accommodate the strongest form of epistemic FPA. We can't do that without embracing something along the lines of the belief condition, and we need to invoke the powers of fictionality to rescue that condition.

Finally, let us address the objection that the belief condition would make womanhood (and other genders) puzzling outliers—for most identity categories, believing that one is a member of that category is not sufficient for being a

---

24. Thanks to Thomas Brouwer for suggesting that I frame this manoeuvre in terms of retconning.

25. Again, is just a rough formulation of Jenkins's account, for simplicity's sake; the fully unpacked version faces the problem outlined in the main text too.

26. Thanks to Rachel Fraser for pressing me to clarify this.

member of that category. A proponent of the belief condition for genders has to explain why gender is different from most if not all other identity categories in this respect.

It's not clear what the explanation would be if genders were real properties of people. But if genders are *fictional* properties of people, we have a straightforward explanation at our disposal. The belief condition is acceptable for womanhood, but not being Asian or being a cat, because the former is a fictional property and the latter are not. Of course, if *races* are fictional properties, the explanation won't be this straightforward. As noted above in Section 3.2, I don't think the motivation I've offered for fictionalism about gender discourse generalizes to racial discourse. But even if fictionalism about racial discourse is true, we would just have to add one further step to our explanation. Given that we don't want to accept an analogue of the belief condition for races, there would be a reason why it wouldn't figure in the best interpretation of that fiction.[27] We would then appeal to that reason in explaining this difference between race and gender.

In short, the constraints on fictional properties concern not what non-fictional properties are like, but rather what makes for the best interpretation of the relevant fiction. The best interpretation of the gender fiction ought to accommodate the strongest form of epistemic FPA, and something along the lines of the belief condition is the only way to do that. The fact that an analogous condition is implausible for non-fictional identity categories is irrelevant. The fictionality of gender can explain why it is consistent to endorse the belief condition in the case of gender (within the scope of the gender fiction) while rejecting analogous conditions for non-fictional identity categories.

4. Pinning Down the Details

Let us now turn to the task of pinning down some of the details of gender fictionalism. One detail concerns what makes gender discourse true in the fiction. In the first subsection, I'll flesh out the notion of an *interpretation* of the gender fiction in order to answer this question. Then, I will turn to the issue of whether we can give necessary and sufficient conditions for womanhood within the gender fiction, and suggest that the best interpretation of the gender fiction is one on which there are two sufficient conditions. Finally, I will consider a major distinction among fictionalisms—hermeneutical vs. revolutionary fictionalism—and argue that gender fictionalism should be a version of revolutionary fictionalism.[28]

---

27. I'll elaborate on the notion of the *best interpretation* of the gender fiction below in Section 4.1.

28. Another major distinction is use vs. meaning fictionalism (see Woodbridge 2005 and Eklund 2011). According to use fictionalism, sentences in the relevant discourse have a literal meaning outside the scope of the pretence, and such sentences are literally untrue. According to

4.1. Interpretations

What makes gender discourse true in the gender fiction? It will be helpful to focus on two different kinds of gender discourse. *Gender attributions* either attribute genders to people, or deny such attributions: for example, 'Maddy is a woman', 'Akira is not pangender'. *Gender theories* are theories that specify the nature of gender and the conditions for being a given gender. To use a rough sketch of Haslanger's (2000) theory as an example: 'Genders are social roles assigned on the basis of perceived sex, and S is a woman just in case she is assigned subordinate social roles on the basis of being perceived as having female reproductive features'.

Of course, this distinction is not exhaustive; for example, 'Women are paid less than men for doing the same work' is an example of gender discourse that is neither a gender attribution nor a gender theory. I'm focusing on gender attributions and gender theories because once we know what makes these kinds of claims true in the fiction, we're in a position to determine the fictional truth value of other kinds of gender discourse. For example, once we know who counts as a woman in the gender fiction, and who counts as a man in the gender fiction, and how much they earn for doing their jobs—a fact about the real world that's imported into the gender fiction[29]—then we're in a position to determine whether 'Women are paid less than men for doing the same work' is true in the gender fiction.

Here are my proposals for what make gender attributions and gender theories true in the gender fiction:

> 'S is a woman' is true in the gender fiction (in context C?) iff according to the best interpretation of the gender fiction (in context C?), S is a woman. (*Mutatis mutandis* for other gender attributions, e.g., 'S is a man', 'S is agender', 'S is not a woman', and so forth.)[30]

> A gender theory is true in the fiction (in context C?) just in case it is the best interpretation of that fiction (in context C?).

Fictionalism can accommodate the possibility that the best interpretation of the gender fiction varies by context. This would yield a version of contextualism

---

meaning fictionalism, sentences in the relevant discourse have meaning only inside the pretence, and outside the pretence they are meaningless. It is plausible that use fictionalism is the default view. Arguing that gender discourse is meaningless, despite robust appearances to the contrary, is a task we should undertake only if there is a very compelling reason to do so—and I don't see one in this case. (Thanks to Mark Balaguer and Paolo Santorio for convincing me that gender meaning fictionalism is unmotivated.)

29. See Walton's Reality Principle (1990: 144–50).

30. Genders shouldn't be regarded as mutually exclusive, in order to accommodate pangender and genderfluid people. (Thanks to Shyane Siriwardena for encouraging me to clarify this.)

about gender terms within a fictionalist framework: for example, which property 'woman' picks out in the gender fiction in a given context depends on which interpretation of the gender fiction is the best relative to that context. For the purposes of this paper, I will remain neutral about whether the contextual parameter is required.

These proposals immediately raise the following questions: what exactly is an interpretation of the gender fiction? And what makes an interpretation the best one?

An interpretation is a kind of Waltonian principle of generation (1990: 38), in that it is a principle for generating fictional truths.[31] Since gender theories purport to tell us what makes it non-fictionally the case that someone is a given gender, it is natural for the gender fictionalist to recast such theories as candidate interpretations of the gender fiction. That is, the various theories of gender should be regarded as competing interpretations of the gender fiction.[32] Since gender fictionalism incorporates every gender theory in this way, it is really a *meta-theory*. It is a theory of gender theories, on which they are characterized as interpretations of a fiction. So it doesn't offer an alternative to existing gender theories; rather, it offers a different way of understanding them. (Compare moral fictionalism—it is a metaethical view, not a first-order moral theory.)

Note that not all interpretations of the gender fiction are *fully fleshed-out* gender theories. That is, interpretations of the gender fiction shouldn't be restricted to explicit, detailed theories offered by people who have spent a lot of time thinking about gender. We should also admit as interpretations the implicit assumptions underpinning the gender discourse of "the folk" (i.e., those not engaged in explicit theorising about gender). Since a folk interpretation is a collection of mostly implicit, unexamined assumptions, it is presumably more likely to be relatively unspecific and to contain hidden contradictions.

One might think that the diversity of theories about the nature of womanhood entails that we're not all talking and thinking about the same thing—for example, that two people who have different beliefs about the nature of womanhood must be attributing different properties to people when they utter sentences of the form 'S is a woman' and believe propositions expressed by them. If this is right, then apparently different beliefs about the nature of womanhood would really be beliefs about different properties (e.g., "womanhood$_1$" vs.

---

31. While I'm inclined to think that gender fictionalism should be elaborated in terms of Walton's framework, I'll leave spelling out the details to future work.

32. There are potentially fruitful connections here with narrative accounts of personal identity; for example, Mark Schroeder argues that someone's personal identity consists in the best interpretation of the totality of their behaviour (2019: Section 5). Exploring these connections here would take us too far afield, so I'll leave this as a task for future research.

"womanhood$_2$"). However, this can't be right. If it were, it would be impossible to disagree about the natures of things. Any apparent difference in beliefs about the nature of a property would amount to beliefs about different properties. But there plainly are such disagreements—disagreements over the nature of womanhood being a case in point.[33]

Now, what makes an interpretation of the gender fiction the *best* one? Plausibly, good-making features of an interpretation of a pretence (in general) include coherence, charitable interpretation of the participants, aesthetic qualities and moral qualities. And to be clear, I'm thinking of good-making/bad-making features of an interpretation in a narrow sense. So, for example, if an evil, omnipotent demon is committed to destroying Earth in the event that the belief condition becomes the dominant interpretation of the gender fiction, this is not a bad-making feature of that interpretation. This is certainly a very bad consequence of this interpretation, but that does not make it bad *qua* interpretation of the gender fiction.[34]

But settling the question of *what makes it the case* that an interpretation is good/better than another/the best is a task that's too big to undertake here. For the purposes of this paper, I just want to argue for a necessary condition on the best interpretation of the gender fiction: an interpretation is the best one only if it accommodates the strongest form of epistemic FPA. The justification for this necessary condition is the argument from Section 2 for the claim that we ought to accommodate the strongest form of epistemic FPA. And given that the belief condition is the only way to accommodate the strongest form of epistemic FPA, it follows that the belief condition is at least a component of the best interpretation of the gender fiction.

The fact that the belief condition is a component of the best interpretation of the gender fiction means that one's beliefs about one's own gender generate fictional truths. This leads to another disanalogy with traditional versions of fictionalism. In the case of moral, mathematical, and modal fictionalisms, the fictional truths are generated by the attitudes and practices of a community considered as a whole. On any plausible way of fleshing out these views, no one person will have the power to make it fictionally true that (e.g.) 2 + 2 = 4, or that killing is wrong, or that there is a possible world in which pigs fly. By contrast, my belief that I am a woman makes it fictionally true that I am a woman, regardless of what anyone else thinks. This is a welcome result for those whose gender identities conflict with the interpretations of the gender fiction that are predominant in their communities.[35]

4.2. *Necessary and Sufficient Conditions*

Note that the belief condition is merely a sufficient condition on womanhood. Might it be a necessary condition as well? There is reason to think it isn't. What about people who do not have any beliefs at all about their genders—for example, a person in a chronically vegetative state (assuming they're not in any mental state that counts as a dispositional belief about their gender)? Or a person raised in a tightly-controlled environment in which they were carefully shielded from all information about gender; somewhat analogously to how Mary (Jackson 1982) was carefully shielded from all colour experiences?

A reason for thinking that such people should count as women is reflected in theories that characterize membership in a gender category in terms of how one is regarded by others. Haslanger's (2000) account, cited above, is an example of such a theory: S is a woman just in case she is assigned subordinate social roles on the basis of being perceived as having female reproductive features. Such accounts are attractive in part because it seems plausible that (e.g.) the quality of medical care that a person in a chronically vegetative state receives will be influenced by the fact that she has the disadvantaged social status of a woman. Similarly, it seems plausible that a person who emerges from the gender analogue of Mary's black-and-white room, baffled by their discriminatory treatment on account of having female reproductive features, counts as a woman. Cases like these constitute reasons to reject the claim that believing that one is a woman is necessary for being one.

That being said, it's not immediately clear how to reconcile this line of thought with the aim of accommodating the strongest form of epistemic FPA. As a first pass, we could say that there is more than one sufficient condition on womanhood, none of which are individually necessary. In particular, we could say that the best interpretation of the gender fiction is one on which S is a woman iff either (i) S believes that she is a woman, or (ii) S is assigned subordinate social roles on the basis of being perceived as having female reproductive features.[36] However, this view as it stands is incompatible with the strongest form of epistemic FPA. For once we have multiple sufficient conditions in the mix, they can yield conflicting results.[37] For example, someone might believe that she is a woman and not a man, but nevertheless satisfy Haslanger's condition for manhood: roughly, being marked as socially privileged on the basis of being perceived as having male reproductive features. On the view as stated, this person

---

36. I'm just using Haslanger's account as a placeholder here; I'm open to a different further sufficient condition.

37. Thanks to Keith Allen for raising this worry.

would count as both a woman and a man, contrary to her belief about her gender—thereby violating the strongest form of epistemic FPA.

In such cases, I think we should say that the best interpretation of the gender fiction positions the belief condition as the most fundamental one. That is: according to the best interpretation of the gender fiction, if a person has beliefs about their gender, then that's the sole determinant of their gender. It's only in cases where someone *doesn't* have beliefs about their gender that a further sufficient condition kicks in. In the case of a Haslangerian further sufficient condition, the best interpretation of the gender fiction is something along the lines of the following: S is a woman iff either (a) S believes that she is a woman, or (b) if S does not have any beliefs about her gender, S is assigned subordinate social roles on the basis of being perceived as having female reproductive features. This is compatible with the strongest form of epistemic FPA within the scope of the fiction; in any case where a person believes that she is a woman and not a man, that belief is fictionally true.

One might object that this is a rather ad hoc manoeuvre; a desperate, inelegant contortion in order to secure the strongest form of epistemic EPA. This would be a worry if we were talking about an account of womanhood offered outside of the fictionalist framework, but remember that we are not aiming to discover joints in social reality here. Our ultimate aim is to secure the strongest form of epistemic FPA, which in turn requires defending the belief condition; and the fictionalist framework offers the most straightforward defence. After all, the reason for adopting the fictionalist framework in the first place is that its flexibility enables us to secure the strongest form of epistemic FPA. So we can and should contort our account of womanhood as much as we have to in order to ensure that result.

4.3. *How We Actually Use Gender Discourse, and How We Should Use It*

An important choice point for a fictionalist view concerns whether we actually assert and believe claims in the discourse at issue. Hermeneutical fictionalism is the view that we don't genuinely assert or believe claims in the relevant discourse. As things already stand, our utterances within the discourse do not aim at truth, but rather are made in a fictional spirit. By contrast, some fictionalisms are revolutionary, in holding that while we actually assert and believe claims within the relevant discourse, we should cease to aim at truth and shift to using it in a fictional spirit (Eklund 2011).

As Stuart Brock observes, this standard characterization of the distinction is silent on a crucial question: who does 'we' and 'our' refer to (Brock 2014: 583)? The more inclusive the referent, the harder it is to defend hermeneutical

fictionalism—for many will insist that it seems to them that they do genuinely assert and believe the relevant claims, and that it doesn't seem to them that they are engaged in pretence (see the arguments outlined in Brock 2014: 579–80). In the case of gender fictionalism, I see no reason to insist that other people are wrong about whether they assert and believe claims expressed in gender discourse and whether they are pretending. The path of least resistance is to adopt a form of revolutionary fictionalism.[38] (Although note that gender fictionalism may well be hermeneutical with respect to those convinced of it, if they continue to use gender discourse—which would be very difficult to avoid.) Moreover, revolutionary fictionalism best reflects the fact that gender fictionalism emerges out of an ameliorative inquiry into the nature of gender—the aim isn't to capture how we actually employ gender terms and concepts, but rather to specify how we should use them given the aim of securing the strongest form of epistemic FPA.

Although I see no reason to insist that people are wrong about whether they're pretending when they use gender discourse, I do want to insist that they are contributing to the continuation and content of the gender fiction without realising it. To motivate the idea that one can contribute to a pretence without pretending (i.e., adopting the attitude of pretence towards its content), let us consider the following case. In recent years, people have started organizing "zombie runs", in which they attempt to increase their speed by pretending that they're running away from zombies. Now, imagine that you're walking down the street, and you come across a group of zombie runners that are exceptionally committed to the pretence. (Suppose they're not even wearing running gear, and they're not breaking character.) Depending on the quality of the runners' acting, you might well be misled into contributing to the pretence; and the runners need not intend for this to happen. Note that *you* are not pretending to run from zombies—you really believe that you are running away from them. However, you are still sustaining the pretence and contributing to its content, for example, by hysterically screaming "I think I see a zombie up ahead!". And you're contributing to the zombie pretence without realising that this is what you're doing.

Just as one can be unintentionally misled into contributing to a zombie run pretence, one can be unintentionally misled into contributing to the gender pretence. On the view I am suggesting, almost all of us have been unintentionally misled; we were unintentionally misled by the adults who raised us, who were unintentionally misled by the adults who raised them, and so on. Of course, while a zombie run pretence has originators who deliberately initiated the pretence, an analogous claim seems implausible in the case of the gender pretence. It's not as if a group of people got together one day and decided to initiate the gender

---

pretence. It's more plausible that the gender pretence organically emerged from a prevalent, mostly implicit hunch that biological differences related to reproduction must make for other kinds of differences. This is extremely speculative; the point here is just that one can contribute to a pretence without pretending.

In short, the idea is that even though hardly anyone is actually pretending that people have genders, the gender fiction is sustained and shaped by everyone's gender-related beliefs and behaviours. And as long as we can make sense of the idea that one can contribute to a pretence without pretending, we need not say that people are unwittingly pretending that people have genders.

The proposal on the table is that we should endorse revolutionary gender fictionalism: that we should stop believing and asserting that people have genders, and shift to pretending that they do. However, there is another issue related to who exactly is included in the 'we'. This issue concerns the distinction between ordinary contexts, and what I'll call 'gender-theoretical' contexts—contexts in which the nature of gender is explicitly at issue, such as a philosophical debate about the metaphysics of gender. In gender-theoretical contexts, it's clear that we should stop believing and asserting that people have genders, because in such contexts we're aiming to get to grips with what gender really is—or, more accurately in the context of an ameliorative project, what it ought to be. So given that gender is best regarded as a fiction, in such contexts we should shift to using gender discourse in a fictional spirit.

Matters are less clear when it comes to ordinary contexts. Note that the reason just given for shifting to using gender discourse in a fictional spirit in gender-theoretical contexts doesn't carry over to ordinary ones. For example, when I say 'I'm a nasty woman and proud of it', I'm just trying to reclaim a sexist phrase frequently used by Donald Trump. I am not at all concerned with the nature of the property of womanhood I'm attributing to myself.

A different kind of reason to shift to using gender discourse in a fictional spirit, one that *would* apply in ordinary contexts, would be any harm that stems from the practice of believing and asserting that people have genders. However, it is not clear that this *general* practice is harmful. Of course, specific instances of such discourse are all too often harmful (e.g., because it's employed in a sexist or transphobic way). But the problem in such a case is with the spirit in which the discourse is employed, rather than with the practice of asserting and believing claims expressed in gender discourse in general. That being said, I certainly haven't said enough to *establish* that the general practice is harmless in principle. Although even if the practice is intrinsically harmful, I am pessimistic about convincing enough people to stop in the foreseeable future, and sceptical about whether it is even psychologically feasible for them to stop (given the centrality of gender to the vast majority of peoples' lives); I'll briefly return to this issue in the final section.

In any case, achieving the aim of gender fictionalism—securing the strongest form of epistemic FPA within the gender fiction—doesn't require that we stop believing and asserting gender claims in ordinary contexts. The strongest form of epistemic FPA is secured within the gender fiction by way of the belief condition, not by the folk embracing the fictionalist metaphysics required to render that condition viable.

Finally, a potential reason *against* a shift to using gender discourse in a fictional spirit in ordinary contexts concerns the main benefit I've argued we get from it (in Section 3.2 above)—namely, its function as "identity glue." It might be that this benefit can be fully realized only if one asserts and believes that one is a given gender. It's not obvious that pretending to be a woman is as good a unifying principle for one's social identity as actually believing oneself to be one.

In summary: it's not clear that there's any reason to shift to using gender discourse in a fictional spirit in ordinary contexts, and there might be a reason *not* to shift. There's much more to be said about these reasons, but for present purposes I just want to briefly sketch how we can accommodate different approaches in ordinary and gender-theoretical contexts if we need to.

We could achieve this by appealing to a distinction among fictionalisms that is more finely-grained than the standard hermeneutical/revolutionary divide (Balaguer 2018). As Balaguer notes, the fact that we assert and believe untrue claims in the relevant discourse as things currently stand doesn't necessarily mean that we should stop doing so—truth is not the only form of success for a discourse. On Balaguer's way of carving up the conceptual terrain, assertional fictionalism is the view that we actually assert and believe untrue claims in the relevant discourse (in contrast to hermeneutical fictionalism); and this view comes in revolutionary and non-revolutionary variants. On a revolutionary variant, we should stop asserting and believing the untrue claims in the relevant discourse and shift to using it in a fictional spirit; on a non-revolutionary variant, there is no need for users of the discourse to change what they're doing. In terms of Balaguer's framework, what I'm proposing is revolutionary assertional fictionalism for gender-theoretical contexts, and non-revolutionary assertional fictionalism for ordinary contexts. That is: people engaged in theorising about gender should stop asserting and believing that people have genders, but people just going about their day-to-day lives can carry on with this practice. This isn't to say that they should carry on with gender discourse as they always have *in all respects*—most importantly, their use of gender discourse should conform to the belief condition, if it didn't already.[39]

One might worry that people are unlikely to conform to the belief condition unless they've bought into the fictionalist metaphysics required to rescue it from

---

the objections discussed above—otherwise, how could they accept it with a clear epistemic conscience? This would mean that respect for the strongest form of epistemic FPA *in practice* would require that people in general take this metaphysics on board, and change their attitudes toward gender discourse accordingly. However, this worry seems to assume that the general public are a lot more like academic philosophers than they actually are. Lots of people outside of academic debates about gender accept the belief condition without these objections coming to mind. (And if they do come to mind, the subject has thereby worried themselves into a gender-theoretical context, and in that context they *should* stop believing and asserting that people have genders.) People can and do accept the belief condition without having the faintest idea of the machinery required to make it *philosophically* acceptable, and general acceptance of the belief condition is sufficient for respecting the strongest form of epistemic FPA in practice.

To summarize, on the proposal under discussion, people in gender-theoretical contexts should stop asserting and believing claims expressed in gender discourse. But as I noted above, revolutionary fictionalism typically comes with an injunction about what to do instead—namely, that we should *pretend* to assert and believe these claims. This injunction seems odd in gender-theoretical contexts, insofar as the main aim of a gender-theoretical context is to ascertain the *non-fictional* truth of the matter regarding the nature of gender. That being said, there may well be points at which engaging in the pretence would be helpful in theorising. For example, a theorist might legitimately draw upon their personal experiences of gender, and in doing so pretend to assert and believe claims expressed in gender discourse (e.g., "As a woman, I think . . . "). We can accommodate both of these observations by saying that it is *permissible* to engage in the gender pretence in theoretical contexts, but not obligatory. That is, instead of saying that we *should* shift to using gender discourse in a fictional spirit in such contexts, we can say that it is permissible to do so. But it is also permissible to refrain from pretending.

In this subsection, I have argued that gender fictionalism should be revolutionary, and that we can accommodate the practice of believing and asserting claims expressed in gender discourse in ordinary contexts if need be. In that case, gender fictionalism should take the form of revolutionary assertional fictionalism in gender-theoretical contexts (with the qualification that engaging in the pretence is permissible rather than obligatory), and non-revolutionary assertional fictionalism in ordinary contexts in which the nature of gender is not at issue.

While there are certainly many more details of gender fictionalism to pin down beyond those just discussed, the view has been specified enough to raise a serious family of objections. If these objections cannot be addressed satisfactorily, pinning down further details may not be worth the effort.

5. The Trivialization Objections

An understandable reaction to gender fictionalism is the worry that it *trivializes* gender—that it fails to take gender seriously enough. This idea can be fleshed out in several ways.

One version of this worry is that gender fictionalism entails that many claims that are central to feminist activism are not true. Such claims include descriptions of what feminists think is wrong with society (e.g., 'women get paid less than men for doing the same jobs'), and prescriptive claims about how society should be changed ('women should be paid the same as men for doing the same jobs'). If the version of gender fictionalism outlined above is true, these claims are not true—because there are no women and men. Doesn't this undermine feminist activism, which is predicated on the truth of such claims?[40]

I think that this particular trivialization worry is relatively easy to address, because feminist activism need not be predicated on the truth of such claims. Although the gender fictionalist has to say that the claims at issue are not true, she can also say that there are true claims in the vicinity: for example, 'the group of people that are women the gender fiction get paid less than the group of people that are men in the gender fiction for doing the same jobs.' Moreover, gender fictionalism doesn't require us to write such cumbersome sentences on our protest signs—we can fight gender-based oppression from within the gender fiction. We can address the serious, non-fictional social problems that have emerged from the gender fiction by using gender discourse in the spirit of make-believe, in order to point towards the true claims in the vicinity.

Another worry is that gender fictionalism cannot do justice to the fact that many of us genuinely value our genders. This is a worry for anti-realism about gender in general, but I don't think it applies to a fictionalist anti-realism. Fictionalism can make sense of valuing one's gender, even though it isn't real: gender figures in a pretence that many of us value.

That being said, there is a further worry that is much more difficult to respond to: characterizing gender in terms of a pretence fails to do justice to the *extent* to which some of us value our genders. I can see why a transgender woman, for instance, who might have gone through hell and back to lay claim to her womanhood, would be deeply offended by the suggestion that she did all that for the sake of a *pretence*. The comparisons above to Star Wars, a fiction intended for mere amusement, unfortunately exacerbate this worry.

However, as Walton insists, "in speaking of 'games' of make-believe we must disavow any implication that they are mere frivolity" (1990: 12). Once we've

---

40. See the representation problem described by Bach (2012: 234). Thanks to Gail Leckie, Rachel Fraser, and Daniel Nolan for independently raising this issue.

appreciated the points the Star Wars comparison was used to make, we can distinguish between relatively *frivolous* fictions and more *serious* ones. I don't have a precise theoretical account of this distinction, but it's intuitive enough. We engage in frivolous fictions primarily for the sake of entertainment. By contrast, serious fictions play more serious roles in our lives—roles like shaping and expressing who we are (one's values, sense of self and community).[41]

One might insist that fictions simply cannot play such serious roles, perhaps based on scepticism about whether "mere" fictions can have substantial real-world effects. But religious texts are counterexamples to this claim. Take, for instance, the Bible, which has indisputably played serious roles in many peoples' lives. It seems clear to most people—including Christians—that the Bible isn't meant to be taken *literally* (at least not in its entirety). Reasonable Christians do not claim (for example) that the Earth really was created about 6000 years ago in seven days, or that Noah actually herded a bunch of animals onto a boat to save them from a global flood.

A disanalogy between religious fictions and the gender fiction is that the former are often regarded as conveying extra-fictive truths, for example, via parables.[42] By contrast, the gender fiction does not tell us anything about what's true outside of it. However, as I argued above in Section 3.2, it performs another important function (providing a unifying principle of our social identities). The reasons why the fictions count as serious are different, but the point of the analogy is just that fictions can play serious roles in people's lives.

Even if religious texts are best construed non-literally, that doesn't mean it is acceptable to disparage or dismiss them. Indeed, since engagement with religious fictions plays such a serious, central role in many peoples' lives, even the non-religious should treat these fictions with respect—out of respect for those who derive value from them. (This doesn't mean that anything goes. For example, one should of course contest an *interpretation* of a serious fiction that is harmful to others.) The gender fiction plays a similarly serious and central role in many peoples' lives. As I suggested in Section 3.2, for many of us, gender is inextricably bound up with our sense of who we are. Thus, the gender fiction should be treated with respect—again, out of respect for those who derive value from participating in it. In short, we should resist the suggestion that construing a discourse as fictional in spirit amounts to trivializing it. Some fictions are profoundly important to many people, and should be respected accordingly.

All that being said, one still might worry that gender fictionalism doesn't make gender valuable *enough*. Surely, one might suggest, it would make sense to

---

41. Recall Bettcher's notion of existential self-identity (2009: 110–12). Even though the goal of this paper is to provide a metaphysical theory of womanhood, Bettcher's notion could be invoked to distinguish serious fictions from relatively frivolous ones.

42. Thanks to Anne Siebels-Peterson for raising this point.

go to great lengths to lay claim to one's gender (as many trans people are forced to do) only if gender is a *real* feature of a person, rather than just a made-up feature in an elaborate pretence. However, such sacrifices are typically rational, simply because most people don't take themselves to be engaged in a pretence when making gender attributions. The fact that one *believes* that gender is real can explain why it is rational to make such sacrifices.

The objector might still insist that if someone *knows* that gender is a pretence, then it wouldn't be rational to make great sacrifices in order to transition—and this consequence (somehow) trivializes decisions to transition. Let's grant for the sake of argument that this consequence would trivialize such decisions. The knowledge that gender is a pretence wouldn't mean it's irrational to make great sacrifices in order to transition. Given that the gender pretence is so deeply entrenched, influencing almost every aspect of our lives, our status within that pretence is of the utmost importance to many of us. So if someone is assigned a status within that pretence that they find unbearable, of course it would be rational for them to transition to another status.

6. The End?

In the course of this paper, I have outlined and defended fictionalism about gender discourse—the idea that talk of gender should be regarded as something like a myth we've been passing down through the generations. This raises the question: should the pretence come to an end? Or should we carry on?

Presumably, it's practically impossible for most individuals to stop; gender is so deeply ingrained in most of us. The live issue is instead whether we should encourage a gradual social shift away from the pretence (by, e.g., refraining from imposing genders on children). I am genuinely unsure about the answer to this question. On the one hand, the availability and historical predominance of bad interpretations of the gender fiction has caused great harm to many people, not just to those in subordinate gender positions (women, non-binary people, trans people), but also those in dominant ones (men, those with genders in the traditional binary, cis people). And I don't see any benefit to engaging in the gender fiction that future humans couldn't get in some other way. On the other hand, the availability and growing acceptance of better interpretations makes me hopeful that it is in principle possible for the pretence to be harmless.

While this question is theoretically interesting, my suspicion is that it is practically irrelevant to anyone alive today. Even if I were to conclude that we ought to work towards the elimination of the pretence, I am not optimistic about convincing enough people to make any progress in that direction in my lifetime. Again, gender is just too deeply ingrained. Our priority in the here and now

should be to keep improving our interpretations of the gender fiction in order to eliminate gender-based oppression.

Acknowledgements

References

Alcoff, Linda (2006). *Visible Identities*. Oxford University Press.

Ásta (2011). The Metaphysics of Sex and Gender. In Charlotte Witt (Ed.), *Feminist Metaphysics* (47–66). Springer.

Ásta (2018). *Categories We Live By*. Oxford University Press.

Bach, Theodore (2012). Gender is a Natural Kind with a Historical Essence. *Ethics*, *122*(2), 231–72.

Balaguer, Mark (2018). Fictionalism in the Philosophy of Mathematics. In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2018 ed.). Retrieved from https://plato.stanford.edu/archives/fall2018/entries/fictionalism-mathematics/

Barnes, Elizabeth (2017). Realism and Social Structure. *Philosophical Studies*, *174*(10), 2417–33.

Barnes, Elizabeth (2020). Gender and Gender Terms. *Noûs*, *54*(3), 704–30.

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* (98–120). Oxford University Press.

Bettcher, Talia Mae (2012). Trans Women and the Meaning of "Woman". In Nicholas Power, Raja Halwani, and Alan Soble (Eds.), *The Philosophy of Sex* (233–50). Rowman and Littlefield.

Bettcher, Talia Mae (2016). Through the Looking Glass: Trans Theory Meets Feminist Philosophy. In Ann Garry, Serene J. Khader, and Alison Stone (Eds.), *The Routledge Companion to Feminist Philosophy* (393–404). Routledge.

Brock, Stuart (2014). The Phenomenological Objection to Fictionalism. *Philosophy and Phenomenological Research*, *88*(3), 574–92.

Burge, Tyler (1979). Individualism and the Mental. *Midwest Studies in Philosophy*, *4*(1), 73–122.

Butler, Judith (1989). *Gender Trouble*. Routledge.

Byrne, Alex (2020). Are Women Adult Human Females? *Philosophical Studies 177*(12), 3783–803.

Dembroff, Robin (2018). Real Talk on the Metaphysics of Gender. *Philosophical Topics*, *46*(2), 21–50.

Dembroff, Robin (2021). Escaping the Natural Attitude About Gender. *Philosophical Studies*, *178*(3), 983–1003.

Díaz-León, Esa (2016). "Woman" as a Politically Significant Term: A Solution to the Puzzle. *Hypatia*, *31*(2), 245–58.

Eklund, Matti (2011). Fictionalism. In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2011 ed.). Retrieved from http://plato.stanford.edu/archives/fall2011/entries/fictionalism/

Hale, Jacob (1996). Are Lesbians Women? *Hypatia*, *11*(2), 94–121.

Haslanger, Sally (2000). Gender and Race: (What) Are They? (What) Do We Want Them to Be? *Noûs*, *34*(1), 31–55.

Haslanger, Sally (2005). What Are We Talking About? The Semantics and Politics of Social Kinds. *Hypatia*, *20*(4), 10–26.

Jackson, Frank (1982). Epiphenomenal Qualia. *Philosophical Quarterly*, *32*(127), 127–36.

Jenkins, Katharine (2016). Amelioriation and Inclusion: Gender Identity and the Concept of Woman. *Ethics*, *126*(2), 394–421.

Jenkins, Katharine (2018). Toward an Account of Gender Identity. *Ergo*, *5*(27), 713–44.

Kapusta, Stephanie (2016). Misgendering and Its Moral Contestability. *Hypatia*, *31*(3), 512–19.

McKitrick, Jennifer (2015). A Dispositional Account of Gender. *Philosophical Studies*, *172*(10), 2575–89.

Mikkola, Mari (2017). Feminist Perspectives on Sex and Gender. In Edward N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2019 ed.). Retrieved from https://plato.stanford.edu/archives/fall2019/entries/feminism-gender/

Reilly-Cooper, Rebecca (2016). Gender Is Not a Spectrum. *Aeon*. https://aeon.co/essays/the-idea-that-gender-is-a-spectrum-is-a-new-gender-prison

Roberts, Amber (2015, July 15). We Talked to the Growing Online Community of "Otherkin": People Who Identify as Non-Human. *Vice*. https://www.vice.com/en_uk/article/mvxgwa/from-dragons-to-foxes-the-otherkin-community-believes-you-can-be-whatever-you-want-to-be

Schroeder, Mark (2019). Persons as Things. In Mark Timmons (Ed.), *Oxford Studies in Normative Ethics* (Vol. 9, 95–115). Oxford University Press.

Stock, Kathleen (2018). What I Believe about Sex and Gender (and What I Don't). https://medium.com/@kathleenstock/what-i-believe-about-sex-and-gender-and-what-i-dont-15da1cba88c6

Stoljar, Natalie (1995). Essence, Identity, and the Concept of Woman. *Philosophical Topics*, *23*(2), 261–93.

Tuvel, Rebecca (2017). In Defense of Transracialism. *Hypatia*, 32(2), 263–78.
Walton, Kendall (1990). *Mimesis as Make-Believe*. Harvard University Press.
Witt, Charlotte (2011). *The Metaphysics of Gender*. Oxford University Press.
Woodbridge, James (2005). Truth as a Pretence. In Mark Kalderon (Ed.), *Fictionalism in Metaphysics* (134–77). Oxford University Press.