Synthese (2024) 204:174
https://doi.org/10.1007/s11229-024-04831-5

ORIGINAL RESEARCH

Solving the inclusion problem: gender without representationalism

Ed Willems[1]

Received: 27 April 2024 / Accepted: 20 November 2024

Abstract
Recent work in the metaphysics of gender mostly focuses on trying to solve the exclusion problem - roughly, the problem of giving a metaphysical account of gender that doesn't exclude anyone from their appropriate gender category. It is acknowledged that no completely satisfactory answer to the exclusion problem has yet been given in the literature; typically such theories fail to account for the diverse experiences and characteristics of trans people. One response is to adopt an anti-realism about gender properties, such as Heather Logue's gender fictionalism. I rebut the move to an anti-realism, showing that it relies on assuming a representationalism about gender vocabulary that we need not hold. I put forward a non-representationalist theory of gender properties that analyses gender vocabulary in terms of its inferential profile, rather than its representational features. This theory constitutes a deflationary realism about gender properties; it solves the inclusion problem, and guarantees the first-person epistemic authority of an individual regarding their own gender identity.

Keywords Gender · Metaphysics · Feminism · Non-representationalism · Deflationism

1 Introduction

The debate over the metaphysics of gender centres around the issue of the inclusion problem: No current theory that purports to outline a definition for gender concepts appears to be able to sort every individual into their *prima facie* correct gender category (Jenkins, 2016, p394). As Katherine Jenkins notes (*ibid.*, pp398-402), this is often because such theories struggle to take into account the diverse experiences and

✉ Ed Willems
ed.willems@york.ac.uk

[1] University of York, York, UK

Published online: 11 December 2024

characteristics of trans people. The inclusion problem thus poses an issue for a trans-inclusive feminism that seeks to find a definition for the concept 'Woman' such that all *prima facie* women are included, in order to pursue feminist projects.

One response to the difficulty with solving the inclusion problem is to deny that gender admits of any such metaphysical definition, and adopt an anti-realist theory of gender vocabulary. Heather Logue (2022) does just this, advocating a fictionalism about gender. Gender fictionalism leaves intact our ordinary uses of gendered language, but without the metaphysical commitments these appear to entail. Logue sees this as desirable, as no other metaphysical account of a property of gender seems able to allow for a strong norm of first-person authority (FPA) about gender avowals, which ought to be a goal of ameliorative inquiry concerning gender (*ibid.* pp132-6). However, there is reason to think that jettisoning metaphysical commitment to genders entails a significant cost. In a 2018 interview, Stephanie Kapusta argues that we ought to try to make sense of the claims of trans people about their gender as bearing metaphysical weight - of the recognition of trans people's gender as a factual recognition, rather than an "ethical" one. Kapusta states that it is only by doing this that we can really validate trans people's claims, noting that "Some sort of radically new approach within analytic social ontology of gender may be required" to do so (Eckstrand & Kapusta, 2018).

The view I offer undertakes this task. It draws on machinery from within existing literature on metasemantics to outline a non-representationalist approach to gender vocabulary. This non-representationalism holds that gender is metaphysically real, but that this does not require us to be able to comprehensively articulate a metaphysical definition for it. This is not to hold that there is nothing informative to say about what having a gender consists in, only that we cannot expect to obtain a complete analysis of gender in a metaphysical key. Within a non-representationalist framework, the way to analyse gender is to reconstruct the language game of gender avowals in order to track the inferential profile of gender terms. On the basis of this language game, we can reconstruct the social significance of gender terms and categories, without an attendant comprehensive metaphysics. Crucially, the inferential structure of the language game guarantees FPA, without thereby committing us to an implausible, contradictory, or exclusionary metaphysical view of gender. Hence the view reconciles the requirement to guarantee FPA with the requirement to solve the inclusion problem.

First, I sketch out the problem and the representationalist move, and argue that it ignores the middle ground position (Sect. 2). Then, I lay out the language game of gender self-ascriptions in terms of the inferential role of a vocabulary of gender and gender identity, showing how this incorporates first person authority (Sects. 3.1 and 3.2). I discuss the importance of the social role of gender concepts and how this can give us a partial insight into the metaphysics of gender and gender identity, and demonstrate how this fits into the account (Sect. 3.3). With the account fully articulated, I show how it solves the inclusion problem (Sect. 4). I then go into further detail on some nuances the view has to allow for, specifically regarding the possibility of lying and being mistaken about one's gender, and the inability to engage competently with the language game of gender self-ascriptions (Sect. 5). Finally, I close out by discussing objections to deflationary views on gender, and how the non-representationalist account solves them (Sect. 6).

2 Non-representationalism

The core discussion in the literature on the metaphysics of gender currently takes the form of trying to solve the inclusion problem: What defining characteristics can be attached to a given gender term such as 'Woman' such that all and only women satisfy it? Social constructivist accounts of gender, as well as others, typically fail to categorise some women - usually some trans women - as women, and thus fail in their categorial aims. This is because, while social constructivist accounts are able to reproduce the ways in which people are ordinarily classed according to gender categories, such classifications mis-categorise certain trans people, e.g.: those who don't pass, or who have not yet come out (*cf.* Jenkins, 2016, p397-9).

The position that seems most able to cope with these challenges is the gender identity (GID) view:

> GID: Any person S falls into a gender category 'G' if and only if S sincerely self-identifies as a G.

The GID view mirrors, to some extent, the conception of gender in trans-accepting circles; this, indeed, is the reason for its apparent ability to correctly categorise trans people, since it apparently provides the only definition sufficiently flexible to accommodate the diversity of trans experiences. It also has the key advantage of according with the norm of first-person authority (FPA) over gender avowals (Bettcher, 2009). The norm of FPA holds that an individual, S, has overriding authority over the beliefs others hold about S's gender. On the pure GID view, peoples' self-ascriptions of gender are taken as authoritative because they metaphysically determine what gender they have.

However, the view faces other problems that mean it is not widely accepted, the most prominent being the circularity objection.[1] The problem notes that self-ascriptions of gender identity are representational - specifically, S's self-identifying as a woman is supposed to consist in her holding a belief that she is a woman. But if GID holds, then S's being a woman simply amounts to her self-identifying as a woman. So the fact represented by her self-identification is her self-identification. This, according to the circularity objection, would make S's self-identifying belief contentless, if not paradoxical, since the ultimate content of S's belief cannot be specified. Therefore, the GID view fails to be a metaphysical account of what having a gender consists in.

Tomas Bogardus (2019, 2022) argues forcefully that the difficulty in accommodating FPA means that *no* account will be able to solve the inclusion problem. Bogardus

[1] Note that Elizabeth Barnes (2022) argues that even the GID view appears to fall to the inclusion problem. Barnes argues that certain cognitively disabled people who are *prima facie* women are, because of their disability, unable to self-identify as women; the GID view thus appears to exclude some women from the category 'Women'. I will discuss Barnes' arguments in detail in Sect. 5.3, below.

points out that there is no property, W, that having the gender identity ‘Woman’ could amount to such that S has W iff they identify as having W (Bogardus, 2022, p1663). This is true even if we take it that W may be indefinable, or that we may be too conceptually impoverished to define it (*ibid*. p1658); it is still the case that S could think that they have W, and yet not actually have it.[2] Thus, whatever underlying property one might take to determine womanhood, there always exists the possibility of being mistaken, hence no account of what womanhood amounts to can satisfy FPA.

One line of reply is to deny that S’s self-ascription of gender aims at representing some inner state that S has. This bears some similarities to the line I want to take, but is not identical - one problem with this line is that it threatens to reject the ability of a person to speak truly about their gender. Heather Logue (2022) has argued that the severity of the inclusion problem means that no satisfactory account of the metaphysics of gender can be given, and that we should therefore be gender fictionalists, that is, that we should regard *all* ascriptions of gender as taking part in an overall fiction of gender. This goes some way to solving the circularity objection. For Logue, the “aim” of the language of gender is not to truly describe individuals as having certain gender properties - there are no such properties for them to have (*ibid*. p139). Instead, the idea of gender properties is a useful fiction, upheld by linguistic practice, which has the function of “[acting] as a *unifying principle* for our social identities” (*ibid*. p141, emphasis in original). However, gender fictionalism has the upshot that, strictly speaking, S’s avowal that she is a woman cannot be true - it is merely counted true within the gender fiction. The fictionalist denies that gender self-ascriptions aim at accurately representing the world, because they take it that there are no gender properties for them to represent. But this is itself a significant cost to the theory, as I argued above.

For this reason, we ought not follow Logue into fictionalism about gender. Logue makes the error of assuming a representationalist metasemantics of gender terms. Following Huw Price, representationalism is the view that for any vocabulary, V, that successfully represents the world, it ought to be possible to give an explanation, in a metaphysical key, of the representational semantic relations pertaining to that vocabulary. For example, if V contains the terms ‘P_1’,… ‘P_n’, which purport to refer to properties, a representationalist account of V would be one that looks for a substantive explanation of what it is about these terms that makes them refer to their subject matter, here the properties P_1,… P_n (Price, 2013, pp8-10 and pp22-6).

The alternative to representationalism is a minimalist account that holds that no substantive, informative account can be found here; the most we can hope for by way of metasemantic explanation is the disquotational sentence “The term ‘P_n’ refers to P_n”– although there will be a pragmatist story to tell that is more detailed (see the following section).[3]

Because representationalism looks for a more substantive account than this disquotational one, it involves employing another vocabulary, besides V, in which the

[2] The limit case here is if W is the property *Thinking one has W*, in which case it faces the circularity objection, although Bogardus does not explicitly address this.

[3] For in-depth discussion about defining representationalism and non-representationalism, see Simpson (2019, 2020) and Tiefensee (2018).

relations underpinning V's semantic properties can be articulated. This typically (although not necessarily) means translating V into naturalistic vocabulary. Where such a translation is difficult or problematic, this creates what Price calls "placement problems", where the purported subject matter of V cannot be 'placed' within an accepted metaphysical picture (*ibid.* pp6-8). Representationalism assumes that if a vocabulary does not admit of a substantive analysis of its representational properties in this way then it does not really successfully represent the world. Hence it is the task of the representationalist to explain why these vocabularies exist, and how they function, given that they have no real subject matter to represent. Fictionalism is one among several responses to this challenge (Price, 2013, p28).

However, the representationalist assumption jumps too quickly from its grounds to its conclusion. It is by no means obvious that every vocabulary V that succeeds at representing should admit of some substantive, informative explanation of its relation to its subject matter. Non-representationalist accounts hold that such a metasemantic link can be inscrutable, and a vocabulary still be in good order, provided we can explicate the rules for its use (Price, 2011, Ch.2). Such accounts are usually deflationist, because they hold that the properties P_1,… P_n are metaphysically real, even though what this reality consists in resists complete explanation. This means that, although these accounts reject *representationalism*, they do not reject *representation*. The terms 'P_1',… 'P'_n still represent their subject matter; they still refer to the properties P_1,… P_n, sentences containing them can still be really true, and so on. Non-representationalist accounts are pragmatist, in that the explanation of the rules of use for the relevant vocabulary looks at the use to which the vocabulary is put, looking to explain its structure with reference to its utility.

I argue that this is the preferred line to take in solving the competing issues of FPA and the inclusion problem for gender vocabulary - reject not representation, as Logue does, but representationalism. I agree with Logue about the extent of the inclusion problem, although I disagree about its cause. A complete, substantive metaphysical explanation of gender properties is impossible, but this does not show that gender properties are not real, only that gender vocabulary does not admit of a representationalist explanation. A non-representationalist, deflationary account of gender terms is therefore the only account that can solve the inclusion problem *and* allow that gender properties are real, gender ascriptions are capable of truth, and so on.

I am not the first to propose a deflationary approach to gender and gender vocabulary. For instance, Louise Antony (2020) offers a deflationary view on which biological sex is the material ground of gender categories, but on which an individual's biological sex does not necessarily determine their gender. Antony argues that gendered social kinds such as 'Woman' and 'Man' are similar to the social kind 'Parent', in that non-social facts give rise to the social kinds themselves. However, non-social facts (e.g.: for 'Parent', being a progenitor of offspring) are neither necessary nor sufficient for someone to fall under the social kind. While gendered social kinds may be unanalysable, discussion of their material ground gives a foothold for feminist projects to discuss gendered injustices. Mari Mikkola (2016, Ch.5) argues similarly that the social kind 'Woman' resists univocal definition, but that this ought not be an obstacle to feminist projects, which can proceed without using it as a basis.

My deflationism agrees with these authors on this basic premise, but holds out more hope that we can delineate a helpful structure for a vocabulary of gender. In this, it is closer to the deflationism of Esa Díaz-León (2018), who sees deflationism as clearing the way for a better understanding of gender properties, rather than necessarily as being a process of shifting the focus away from questions about these properties. Where my account differs is in locating the problem specifically in the representationalist assumption, hence 'non-representationalism'. The account also goes further than that of Díaz-León in using non-representationalist machinery to delineate a language game of gender avowals that meets the requirements of an inclusive gender vocabulary (Sect. 3, below). This is to say, where other deflationary accounts attempt to *dissolve* the inclusion problem, the present account attempts to use deflationist machinery to offer a solution.

Accounts of gender(/identity/vocabulary) typically note whether they are attempts at hermeneutical inquiry or ameliorative inquiry, or some other branch. As the names suggest, hermeneutical inquiry aims to cash out extant concepts of gender, while ameliorative inquiry aims for the concepts we *ought* to have if we want to achieve the aim of maximising social justice. The present account fits broadly into the latter category, since it is a form of deflationism. Deflationary theories are, of necessity, motivated on pragmatic grounds, where this is understood to include practical utility of the language form for facilitating our goals and projects. Hence they are intrinsically geared towards ameliorative inquiry, if we take these goals to include social justice, as we should.

There may be a worry that the non-representationalist account presented here cannot be a form of ameliorative inquiry, because by its own admission it is unable to substantively define a target concept of gender. I take this worry seriously, but I believe that the account ought to be considered a form of ameliorative inquiry. First, as above, the account is specifically in the business of improving on our conceptual-linguistic practices in order to advance the cause of social justice. This places it under ameliorative inquiry on its broadest construal, as the project of "[elucidating] "our" legitimate purposes, and what concept of F-ness (if any) would serve them best" (Haslanger, 2012, p376, scare quotes and round parentheses in the original).

There are several possible ways of elucidating concepts, the deflationist methods canvassed below being among them. Although the present account does not offer a representationalist specification of the content of its target concepts, this is not the only way of specifying that content. An alternative is that one could learn the language game of applying gender terms by learning the rules articulated in Sect. 3 below. The content of the target concepts can then be specified from within the target language (this process has the same relevant structure as Robin Dembroff's "imitation" approach - Dembroff, 2018). I will discuss further this process of specification in Sects. 3.2 and 3.3 below.[4]

The criteria for success for the non-representationalist account, then, are as follows: First, the account must solve the inclusion problem. I take this to mean that the

[4] For discussion in the literature on difficulties with and strategies for specifying target concepts for ameliorative inquiry, see also Mikkola (2009), Bogardus (2019) and Díaz-León (2020). My thanks to an anonymous reviewer for raising the concern regarding ameliorative inquiry.

account must classify everyone with the gender identity ‘G’ as a G. Included in this is a requirement that the account attribute metaphysical reality to gender properties. Next, the account must not fall foul of the circularity worry, i.e.: the content of a claim that one is gender G must not be that one *claims to be* gender G. And, finally, the account must guarantee first-personal authority over gender self-ascriptions.

An account that can do all of these things will satisfy the demands of ameliorative inquiry, first, by providing an understanding of the term ‘Woman’ suitable to underpin trans-inclusive feminist projects, and second, by validating the epistemic authority of trans people concerning their own genders, itself an issue for social justice.

3 The account

Non-representationalist analysis of a vocabulary proceeds by analysing the inferential rules governing the vocabulary, in order to sketch the structure of the relevant language game. It does this by charting the introduction and elimination rules - when we are entitled to use a term, and what language game moves it entitles us to, respectively - in order to establish a term’s inferential profile. Michael Williams sketches a formula for such analysis, which he calls an “explanation of meaning in terms of use”, or “EMU” (Williams, 2013, 2015). The EMU contains separate “clauses” for the inferential rules for a term, the epistemic features of those rules, and the pragmatic utility of a language game with that structure. I will follow roughly Williams’ pattern, with a slight adjustment to the format, explaining at each juncture why the relevant aspect of the language game works as it does.

3.1 Inferential model

What should the inferential rules - the introduction and elimination rules - for ascribing a gender term such as ‘Woman’ be? I propose that they should follow the model Wittgenstein (1950) gives for first-personal avowals of internal states. I am not the first to draw this connection between Wittgenstein’s expressivist approach and FPA (*cf.* Bettcher, 2009, p99). However, the approach of taking this expressivism to underpin a metasemantics of gender terms - i.e.: a non-representationalism about gender vocabulary - has not hitherto been undertaken in the literature. This is the approach that I propose. Modelling gender terms in this way primarily means giving an inferentialist semantics for instances of gender self-identification. Such inference rules track inferential entitlement - which inferences from which premises are legitimate - and thereby track epistemic warrant.

I will use the term ‘Woman’ in the following as an example, but the explanation is intended to be schematic, so as to apply to all other gender terms. For clarity, for the purposes of these inference rules, I take sincere self-identification to encompass speech acts (broadly construed) declaring one’s gender. To start with, we have a sufficiency condition:

> Inf(1): If one has warrant to believe that S sincerely self-identifies as a woman, then one has warrant to believe that S has the gender identity ‘Woman’.

The next rule establishes the necessity that holds between self-identification and warranted beliefs about gender:

> Inf(2): One has warrant to believe that S has a gender identity other than 'Woman' only if one has warrant to believe that S does not sincerely self-identify as a woman.

Inf(2) is meant to guarantee the strong connection between sincere self–identification and what one is warranted to believe about another's gender. In combination with Inf(1), it provides a picture on which someone's sincere self-identification always trumps opinions to the contrary. This guarantees FPA for individuals regarding their own gender, as I will discuss further in Sect. 4. I discuss cases where individuals' self-identifications might be considered problematic in Sect. 5.

3.2 Metaphysical clause

Note that I have not, at this point, given an answer on when S is a woman, metaphysically speaking. Instead, the above is a set of usage rules for the term 'Woman'. For the non-representationalist, this is the basis to work from to construct an informative account of the concept. Because the non-representationalist adheres to a deflationism about semantic representation, the inferential section of the EMU is semantically minimalist, since, by deflationist lights, it is not possible to provide a complete, substantive unpacking of the conditions under which the analysed vocabulary applies. This means that the explanation often falls back on disquotation as a means of providing truth conditions for the vocabulary in a minimalist way (Williams, 2013, pp143-150).

For present purposes, the "metaphysical clause" (my appellation) forms part of the inference rules for gender vocabulary. The metaphysical clause for 'Woman' on the present account is as follows:

> Inf(3): The sentence "S is a woman" is true if and only if S has the gender identity 'Woman'.

Something that will be immediately clear is that the term 'Woman' appears on both sides of the biconditional. This, again, is a feature of the account's semantic minimalism, and as such is not problematic. The same term can appear in both the explanans and the explanandum without implying problematic circularity, because the project here is not to point to the subtending properties that the term represents, but to specify its truth conditions by whatever means - as Williams notes, this feature implies no "representationalist backsliding" (*ibid.* p141).

It follows that the metaphysical clause of an EMU is likely not to be very informative, but this is likewise not problematic. The clause can be interpreted by anyone who has mastered the vocabulary with which the explanans is articulated. That this is the same vocabulary used in the explanandum is not problematic unless we adopt the representationalist principle that it should be able to be articulated in some *other* vocabulary. Again, the present account rejects this very principle. Instead, the non-representationalist EMU points to the structure of the language game in which the vocabulary is situated, to inform about how it is used.

Inf(3) does have a metaphysical upshot: The facts about an individual's gender supervene on their gender identity. Hence the non-representationalism put forward here properly falls into the category of "GID (Gender Identity)-first" views (Rowland, 2023a, pp802-4). The reason for this is simply that no other sort of view is capable of solving the inclusion problem. What distinguishes it from other GID-first views is that it maintains that we ought not require a complete, informative definition of gender identity to be given in another vocabulary, which is how it solves the subsequent circularity problem that afflicts those views.

The worry may persist that it is obscure exactly what the content of gender terms is, on this view, since we have not, for instance, explained what it is that differentiates different gender identities. The account leaves open a couple of ways to specify this content. First, following Robin Dembroff (2018), we can learn to use the trans-inclusive vocabulary largely independent of the ability to state reductive definitions for its terms, by imitating the usage of those who have already mastered it, until we ourselves come to internalise and intuit it. This process of coming to understand "how to go on" is par for the course for learning a new language game, as will be familiar from Wittgenstein (1950). In terms of representation, it is a process of developing sensitivity to the different properties tracked by the terms of the gender vocabulary, such that one can reliably apply them accurately. The animating idea of non-representationalism is simply that representationalist analysis is unequipped to capture these properties fully, but that there are nevertheless other means of specifying them. Thus we can both articulate the inference rules for applying gender vocabulary in ordinary contexts (Inf(1) and Inf(2)), and allow that people can, through habituation, learn to recognise the properties to which gender terms refer, without the need to subject gender vocabulary to representationalist analysis.

Despite the rejection of representationalist analysis, gender (identity) is not completely inscrutable on the account given above; the non-representationalist can still offer some explanation of the features of a property of gender identity. As will be made clear in the pragmatic clause, below, the most appropriate way to think about gender identity is in terms of a relation to social norms and structures. This is not the same as a straightforward social constructivist theory of gender (e.g.: Ásta, 2018, Chapters. 3&4, Haslanger, 2000), although it takes on board many of their points. On the present theory, gender is what I will call a "para-social" property, that is, a property that is determined by a relation (gender identity) to a set of socially-constructed norms. This bears obvious similarity to Katharine Jenkins' (2018) view (another GID-first view) of gender identity as a gendered social "map". However, it is not identical with that view; strictly, Jenkins' view is reductionist, since it reduces gender identity to a set of attitudes and dispositions, hence it falls foul of the inclusion problem (see Barnes, 2019, p8 and Logue, 2022, p146).

Non-representationalism avoids any reduction of gender identity, so it avoids this pitfall. Thinking of gender identity as a para-social property, we can state that S's having the gender identity 'Woman' consists in her bearing a relation to the social norms and structures pertaining to women. What we cannot do is give the set of characteristics S must have in order to to count as satisfying this relation in respect of the gender identity ‘Woman’. I will discuss further the notion of gender identity as a para-social property, as well as Jenkins’ view, in the following section.

Note that the inferential profile of gender terms here captures the metasemantics of avowals of both binary and non-binary gender-identities, including those of agender people. In the following, I treat being agender as having an identity that precludes classification within a gender category, per the phrasing of Inf(2). I remain uncommitted on whether ‘Agender’ is itself a gender identity or a lack of such, as this makes no difference to the view’s ability to correctly categorise agender people.

3.3 Pragmatic clause

Williams’ format for the EMU reserves a space for a pragmatic explanation of the use of the target vocabulary. This serves a few functions, chief among which is to make clear the things the vocabulary allows us to do that are desirable, which we couldn’t do without it, and that other versions of the vocabulary, with different rules, wouldn’t be able to deliver (Williams, 2013, p135). One thing this achieves is showing why this version of the vocabulary is desirable; in a deflationist context, we may have the option to abandon the target vocabulary, e.g.: for another better suited to our aims, and the pragmatic clause makes the case for retaining this version of the vocabulary (remembering that, in the deflationist context, pragmatic utility is the only viable way to settle these questions). A side benefit of the clause is that it helps further illustrate the structure of the language game in a way that stating the inferential profile of the vocabulary doesn’t.

Initially, it seems difficult to point to the utility of gender vocabulary in the same way non-representationalists point to e.g.: the utility of talk of truth or reference. Gender norms are, famously, largely in the business of repression. When we get rid of the gender norms whose societal function we find unjust or reprehensible, it’s not immediately clear what the function is of those norms that remain. A Victorian might have argued the pragmatic utility of a norm restricting the voting rights of women on the basis of their being less rational than men; with a more egalitarian attitude towards different genders, it becomes more difficult to point to clear-cut, practical reasons for treating them differently.

One reason we might point to, albeit not immediately a practical one, has to do with self-expression. It is of benefit to individuals to be able to express themselves in relation to gender norms. This expression can take the form of gender presentation, but it can also take the form of simply interacting with these norms in very ordinary ways. Such interaction, especially for trans people, can be very affirming, just as being unable to interact in that way can be psychologically damaging, leading to the experience of gender dysphoria and its myriad comorbid symptoms (Kapusta, 2016; Kirkland, 2018; Ritchie, 2021).

A growing clutch of theories propose to understand aspects of gender as a relation of an individual to gender norms. Of immediate interest here is Katharine Jenkins’ (2016, 2018) norm-relevancy account of gender identity. For Jenkins, to have the gender identity ‘G’ is to experience the norms pertaining to Gs as being relevant to oneself. Jenkins links this to having an internal “map” for behaviour in different social situations, where one’s map determines one’s gender identity, because of the differences between maps for different genders (Jenkins, 2016). The norm relevancy point is useful to us here because it seems to point to an origin for the feelings of discomfort experienced by trans people (and, indeed, by anyone) when they are held accountable to norms that they feel ought to be irrelevant to them, e.g.: when a trans woman is taken to violate male norms of dress, to which she ought not be beholden. Such treatment constitutes social disenfranchisement, i.e.: depriving an individual of the ability to elicit recognition of their social intentions (Jenkins, 2018, p732, citing Lugones, 2003).

Similarly, Rowland offers a "fitting treatment" account of gender judgements (Rowland, 2023b). Rowland's account states that to judge someone to be a gender G is to judge it to be fitting to apply the norms for Gs to them. "Fitting" here is being used in a very specific way. Rowland states that it does not mean "Morally fitting, all things considered." For instance, if a demon insisted that you admire it, or else it would kill your family, it would be morally fitting, all things considered, to admire the demon in order to save your family. However, it would not be fitting in the sense that it is fitting to admire only admirable things, because the demon is not admirable; it is this sense that Rowland has in mind (*ibid*. p4). Rowland makes clear, drawing on trans people's accounts, that great psychological significance can attach to this sense of fittingness; being judged by norms that don't fit can be psychologically harmful, and conversely, simply having others judge that it is fitting to apply certain norms to oneself can be psychologically liberating and fulfilling (*ibid*. pp9-10). As Rowland notes, trans people will put up with great costs, including social costs such as ostracisation, to achieve this sort of fittingness in the judgements of others (*ibid*. p6).

The particular norms relevant to each gender category need to be specifiable independently, at least to a degree. The phrase "norms pertaining to women" has a familiar ring of circularity, since we are partly specifying who 'women' refers to based on a relation to these norms. In fact, there is no circularity here; the phrase "norms pertaining to women" used above is intended to refer simply to norms of femininity or womanhood - for instance norms pertaining to feminine presentation, behaviour and so on. I take it that this reading, which is not ontologically committing, is what Jenkins and Rowland use to specify the relevant norms.

This also allows us to distinguish different gender identities by distinguishing between norms for different gender categories. That is, although the inferential structures detailed in Inf(1)-Inf(3) are the same with respect to both 'Woman' and 'Man', the norms relevant to each gender identity are different. Hence in inferring via Inf(1)-Inf(3) that S is a woman, we infer that S has the gender identity 'Woman', that is, that S bears a certain relation to social norms of womanhood that makes it appropriate to judge S by them. If we were to infer that S was a man, that inference would be structurally similar, but would have different consequences, according to the different norms to which S bears a relation, and by which it is appropriate to judge them. Thus we can explain the different contents of the two judgements.

If the worry about specification persists, the non-representationalist can draw on what Louise Antony calls the 'material ground' of gender norms (Antony, 2020). Antony argues that the sexual dimorphism of the human species is what originally gave rise to gender norms, and thereby concepts of gender. Specifically, gender norms historically originated as a system of control and exploitation of the reproductive powers of biological females (*ibid.* p533). This is not, as Antony is at pains to point out, to say that gender amounts to sexual biology. For one thing, biological sex itself is not as clear cut and dimorphic as the norms make it out to be (*ibid.* p534). To talk about the material ground of a concept 'Woman' is not to outline its extension, per Antony's deflationism. For present purposes, though, the concept of material ground should suffice to specify the relevant norms. For example, the phrase "the norms pertaining to women" can be taken to refer to the norms that have as their material ground the female role in reproduction. That trans women, for instance, do not occupy this reproductive role is not relevant; trans women have the gender identity 'Woman', which consists in a para-social relation to the relevant norms, which are individuated by their historical origin.

I take it that this framework can straightforwardly be built outward to account for non-binary gender identities in terms of the norms that have the male and female reproductive roles as their material ground. For example, a genderfluid person might experience felt-relevancy, in Jenkins' sense, with respect to different norms with different material grounds at different times, while an agender person will not experience any of the norms in question as being properly relevant to them, and so on. Again, recall that this is not to provide a reductionist account of gender identity, but to flesh out what we can informatively say about gender as a para-social property, despite its resisting a full representationalist analysis.

Jenkins' and Rowland's accounts of felt-relevancy and fit are continuous with each other, and are largely continuous with the account given here. Importantly, they offer perhaps the most informative picture available on what it is to have a gender identity. The important thing for the pragmatic clause of the non-representationalist account is that they make clear the sense in which the inferential profile for gender vocabulary detailed above is socially significant, and psychologically significant to individuals.

But the non-representationalist account can do something for "fit/relevancy" accounts also. The accounts do not make it clear exactly where the phenomena of fit and relevancy arise from. We can see this most clearly with relevancy: Jenkins states that gender identity consists in the felt relevancy of gender norms to oneself, but does not explain why people feel the pull of gender norms differently to one another - some so strongly that they undertake great burdens to observe the norms they feel are relevant to them. Jenkins regards it as a desideratum of an account of this type that it *render plausible* the idea that gender identity is worthy of respect, though not that it *explain why* this is the case (Jenkins, 2018, p731). As noted, Jenkins argues that her account renders respect for gender identity plausible on the basis of social disenfranchisement of those who pursue felt relevancy, but this adverts to the strength of felt relevancy, rather than explaining it. An explanation of the origin of felt relevancy is a genuine concern; a trans-exclusionist might argue, for instance, that the strong feelings of norm-relevancy experienced by trans people arise from pathology, and are thus not to be respected, despite their cognitive consequences. Similarly, for fittingness, Rowland gestures towards different cases to classify different types of fittingness, but does not give a definition. This leaves open the question whether trans-inclusive and -exclusive communities are using the same definition, and thus whether they can meaningfully disagree.[5]

The non-representationalist account recontextualises fit and relevancy as facets of the gender language game. On the account, to judge that S is a woman involves judging that it is fitting to apply the norms for women to S. This is simply to observe the language game interacting with our everyday conduct - what Wittgenstein might call a “way of life”. To judge that S *fits* the concept ‘Woman’ is just to judge that, in order to play the gender language game properly, one has to attribute womanhood to S, and hold her to the pertinent norms. Likewise, bare relevancy can simply be understood in terms of extension: The norms for women are relevant to S because S is a woman. The phenomenon of *felt*-relevancy that is Jenkins’ focus is thus the perception by an individual that it is appropriate for them to interact with this way of life in a certain way, as bound by the norms governing a certain category of people. This is not *constitutive* of gender, on the non-representationalist account, but rather a corollary that follows from the way the language game is constructed. Fittingness and (felt) relevancy ought to be respected because we do an epistemic injustice to individuals if we do not respect them, i.e.: by breaking the rules of the language game we cheat those individuals of their due acknowledgement.[6]

We can use these accounts, then, to flesh out what we mean by gender identity, by pointing to the role that the concept has with regard to social norms. Absent other considerations (per the inferential clauses), one can use the concepts of fittingness and norm-relevancy as an aid to assessing gender identity. Likewise, other accounts of the social significance of gender properties, such as Haslanger’s (2000, 2017) can be informative in this regard. That said, no such account is to be taken, for present purposes, to be a *reduction* of gender identity. The essences of particular gender identities still resist complete metaphysical explanation, though, importantly, not to the degree that we cannot outline a functioning language game that preserves their social importance (and guarantees inclusivity, FPA and so on). This is not spooky or undesirable - gender is just one of those concepts (and there are potentially many) that resist cashing out in terms of a set of corresponding essential features, that is to say, that resist representationalist explanation. The non-representationalist account explains why it is that the phenomena of fittingness and norm-relevancy cannot be justified in an analogous way to other, less inscrutable phenomena. It is not the case that felt relevancy arises from the imaginations of trans people, or fittingness from those of trans-inclusive epistemic agents. Rather, these phenomena are manifestations of linguistic competence with respect to a vocabulary whose subject-matter resists representationalist explanation.

[5] Note that deflationism takes this problem in its stride; according to the deflationist, many (or all) metaphysical disagreements arise from talking at cross purposes using different languages. The appropriate response is thus to take seriously the choice of which language, and thus which conceptual structures, to adopt.

[6] For a detailed discussion of the effects of testimonial injustice with respect to trans people, see Fricker and Jenkins (2017).

4 Solving the inclusion problem

The conceptual outline of a non-representationalist gender vocabulary given above answers the need of a trans-inclusive feminism to provide a conception of womanhood that encompasses all (and only) women. That is to say: It solves the inclusion problem. For present purposes, we can view the goal as being to answer the question "Who is a woman?" in two senses - one metaphysical, the other epistemic. To the metaphysical question "Who is a woman?/Who has the property of womanhood?", the answer is simple, and deflationary: All and only those people who have the gender identity 'Woman' are women.

It might seem as though some trick has been pulled here - we have not given a definition of gender identity, so how can a definition of gender properties in terms of gender identity tell us who is included in the goals of feminism? But there is no sleight of hand going on - this is part and parcel of the deflationary approach.

One place the worry could potentially stem from is the representationalist presupposition that only those concepts that admit of reductionist definitions can be used to discern real categories and real properties. Deflationists (and others) have argued for years that this is not the case (e.g.: Carnap, 1950, Boghossian, 1990, 2002, Price, 2011, Ch. 6, Zalabardo, 2019, to name a few), and the non-representationalist account I have given openly rejects this principle. If any property was ever a candidate for being representationally and metaphysically inscrutable, it is gender, as successive attempts at definition have shown. But this comparative inscrutability does not imply its unreality; anti-realist theories, such as fictionalism, make that leap too fast. Nor does it imply that we can say nothing informative about gender identity; the account of felt relevancy and fittingness canvassed above show that there are informative features of gender identity, construed as a para-social property, that we can analyse even within the restrictions imposed by a non-representationalist framework. On the non-representationalist account, gender is both real and, it turns out, readily assessable in most cases.

The other reading of the question is epistemic: How do we know who to count as a woman, in the course of everyday life, and who not to? Given that we haven't presented a complete account of what having the gender identity 'Woman' *consists in* (and, indeed, have maintained that such can't be done), the worry might well persist that we haven't yet cleared things up for feminism. How are we to assess, given the difficulty with defining gender identity, whether a given person, S, is a target for feminism - whether it is the project of feminism to guarantee their rights, or not?

The non-representationalist account provides a reasonable guide to how to make this assessment: First, consult S's own avowals, since, if such are available, this is the only way to obtain rationally justified beliefs about their gender identity. If no such avowals are available (for example, because S lacks competence in the language game required to make such avowals - see Sect. 5.3 below), one should employ intuitions concerning fittingness, what one understands about what gender norms (if any) S feels they ought to be judged by, and so on, to form an opinion.

Note that this applies to attempts to discern one's own gender identity via introspection: One ought to consult one's own feelings of felt-relevancy of gender norms to oneself, if any, in order to assess what gender category one takes to "fit" oneself.

Again, the non-representationalist account here occupies a sort of middle ground. Because of their broadly Wittgensteinian approach, the non-representationalist cannot offer a comprehensive epistemological or inferential framework for performing this sort of introspection. However, they can offer an informative (if partial) account of what gender identity consists in to help guide it (Jenkins, 2018; Rowland, 2023b), as well as a blueprint for how to develop a sensitivity to the properties to which gender terms refer (Dembroff, 2018). Using this understanding, one can form an opinion through introspection about one's own gender identity.

The non-representationalist account provided above avoids the circularity worry because it does not tie the facts about one's gender to one's avowals; rather, it holds the facts about one's gender to be determined by one's gender identity. Epistemic warrant, not gender identity itself, is then contingent on avowals. Per Bogardus' (2022) argument, the view allows that there is some property, W, in virtue of which one counts as a woman, and that one could have or fail to have it irrespective of what one thinks, one's gender avowals etc. However, it neither falls to the inclusion problem, nor forfeits FPA.

To see why, note that there is an ambiguity here on which Bogardus appears to trade, namely that between "self-identifying" meaning *avowing a gender identity*, and meaning *having a gender identity*. The non-representationalist view distinguishes these. First, it solves the inclusion problem because everyone who has the gender identity 'Woman' has the property W, metaphysically speaking. Hence there are no women we exclude from womanhood. Second, since one cannot have warrant to believe that S does not have the gender identity 'Woman' (and therefore the property W) in the face of S's (sincere, competent) avowals to the contrary, the view ensures that FPA is respected.

The only remaining problem would arise if we took it as a desideratum that self-identifying as a woman, in the sense of *avowing the gender identity 'Woman'*, be sufficient for having W. But this is just to lock the theory into implausible circularity. For one thing, it takes the "authority" of FPA in the wrong sense; the norm of FPA guarantees epistemic authority, but does not establish that one's avowals metaphysically determine one's gender (Bettcher, 2009, p111). For another, it is actually at odds with trans narratives of self-discovery: Trans people often attest to having discovered their true gender, contrary to previous sincere avowals, and hence to having been mistaken previously (Logue, 2022, p130). The ideal theory would be able to model these narratives, which it can't do if it makes gender metaphysically dependent on avowals. Again, the present theory can accommodate these narratives, all while ensuring that FPA is respected. The problem arises only if we equivocate the two senses of "self-identify" in outlining FPA.

The view given above is broadly inferentialist: It reconstructs the language game of gender avowals in terms of the inferential moves individual speakers are licensed to make. It constitutes a neo-expressivism about first-personal avowals, in that it takes first-personal avowals not to admit of explanation in terms of representing inner states, but nevertheless takes them to be properly thought of as assertions (Bar-On, 2004, Ch.10; Fan, 2022). By inferentialist lights, for a sentence to convey assertoric content is for it to factor into what Robert Brandom calls "deontic scorekeeing" - the practice of tracking a speaker's inferential commitments and entitlements (Brandom, 1994, *passim.*). Gender ascriptions do this, on the view offered above, hence they can be used by an individual to make factual assertions about their own gender, and by others as the basis for inferences about such.

For the non-representationalist, first-person avowals made by S license certain inferences about S's internal states. What is special about first-personal avowals is that, concerns about problem cases aside, if S makes such avowals available, these constitute the only authority sufficient to provide this license, as follows naturally from Inf(1) and Inf(2). Thus the neo-expressivist view on first-personal avowals satisfies FPA, because in expressing them one makes an assertion that is not normally subject to contravention by others.

5 Problem cases

In this section, I will discuss some important problem cases for any account of gender and gender terms that attempts to account for FPA, namely cases of deception, cases where an individual is mistaken about their gender identity, and cases where an individual lacks competence in the gender language game. Theories of first-person avowals generally ought to account for such cases. I hope that the approach to these cases will show that the present account is robust and able to handle the nuances of different situations regarding gender avowals in a plausible way.

5.1 Mistakes

The requirement to guarantee a strong norm of FPA might seem to be in tension with the possibility of being mistaken about one's own gender. However, since the non-representationalist theory separates the metaphysical question of which gender identity (and therefore gender) one has, and the epistemic question of what we are licensed to believe about someone's gender identity, it solves the tension quite naturally. As noted above, archetypal trans narratives have it that one can be mistaken about one's own gender identity - introspection is an epistemic process, and not an infallible one. However, FPA does not require that one be an infallible knower of one's own gender identity, only that we ought to respect individuals' sincere avowals of their own gender identity as authoritative.

Talia-Mae Bettcher (2009) argues that introspection is so fallible that it does not provide a strong rational basis for a norm of FPA. Here, the non-representationalist is actually committed to agreeing; cashing out the normative force of FPA in terms of the reliability of introspection would be deriving its force from representational accuracy, which would be to lapse back into representationalism. For the non-representationalist, the force of FPA derives from the structure of the language game, which is justified pragmatically on the basis of its utility and positive social outcomes. This is analogous to the case of avowals of sensations in Wittgenstein (1950): People are not infallible concerning avowals of pain, but nor is the authoritative force of such avowals derived from introspective accuracy. Rather, it is simply a feature of the structure of linguistic practice that if someone sincerely asserts "I am in pain," it is not rational to doubt them under normal circumstances. "Normal circumstances" here are simply circumstances where the hearer has no reason to doubt that the report is sincere, and that the speaker is competent in the language being used - where the language game is being played competently and in good faith, as it were. A language game of pain avowals that allowed that a hearer could rationally question such avowals, under such normal circumstances, would be unworkable, or otherwise undesirable. Thus the structure of the language game - as supporting FPA, even though mistakes are possible - is supported pragmatically.

Applying this structure to a language game of gender avowals likewise delivers FPA regarding gender identity, without the need to make FPA contingent on the representational accuracy of introspection. This structure of the language game is therefore justified by the project of ameliorative inquiry, since a norm of FPA is justified on ameliorative grounds, and only such a language game can guarantee FPA. Thus, analogously with pain reports, when a subject S says "I am a woman", unless we have good reason to doubt S's sincerity or competence (see the following sections), we do not have good reason to doubt S's avowal. This applies even if it later turns out that S was wrong, e.g.: because S previously took themselves to be a cis woman, and later realised they were a trans man. In such a case, the belief "S is a woman" will turn out to have been false but justified at the time. This is what we should expect if we want a theory that upholds FPA, but doesn't require implausible infallibility; as rational belief-formers, we can do no better than to trust S's avowals, even though they may turn out ultimately to be mistaken.

5.2 Deception

People sometimes make disingenuous claims about their gender identity. As regards detecting deception, the non-representationalist has to walk quite a fine line between, on the one hand, providing a robust epistemic framework that doesn't require implausible credulity, and on the other, avoiding any erosion of FPA by taking emphasis away from individuals' gender avowals. I believe this can be done by placing emphasis on sincerity, as that word is used in Inf(1) and Inf(2) above. Specifically, for an individual S who avows identity 'G', we ought to differentiate the suspicion that *S is not being sincere* from the suspicion that *S does not have gender identity 'G'*. The former can be warranted if we have independent reason to believe S intends to deceive, hence that they are not sincere in their avowal. It can thus act as a legitimate defeater for the knowledge that S has gender identity 'G'. The latter does not provide this warrant, as S's FPA rationally ought to override suspicions that concern their gender identity directly, per Inf(1) and Inf(2).

To illustrate, consider the following case: Suppose A is a trans woman, who hides her real gender identity for fear of discrimination, thus presenting as a man. We might imagine conceiving suspicions from A's behaviour, personality traits etc. that A does not have the male gender identity that she avows. The question is whether these suspicions - or beliefs corresponding to them - would be rationally justified.

In the case described, it does not look as though they can be. Assuming A's competence, in order for us to rationally doubt A's avowals, we would have to have reason to doubt her sincerity, per Inf(2). However, the only reason to doubt her sincerity in this case comes from suspecting that she has a gender identity other than the one she avows. That is, the reason we suspect A might not be sincere is because we already suspect she may have the gender identity ‘Woman’. But this is the wrong way around; according to Inf(2), in order for the belief that A has the gender identity ‘Woman’ to be rationally justified in the face of A’s avowals to the contrary, we would already have to have justified reason to believe A to be insincere. To doubt A in this case would be to reason in a circle: The belief about A’s gender identity is needed to justify the belief that A is insincere, but the belief that A is insincere is needed to justify the belief about A’s gender identity. Because there is no independent reason to doubt A’s sincerity, we cannot justifiably doubt her avowals.

Are there cases where there is independent reason to doubt someone’s avowals? There are examples to be found, although they are few in number (as they should be, since by its nature FPA ought to be assumed in most cases). For instance, suppose that B is a known anti-trans activist. B is a cis man, but claims to be a trans woman in order to access women-only spaces, to protest the admission of trans women into these spaces. Since B’s deception is intentionally transparent, it would count against the language game of gender avowals as reconstructed here if it mandated that we believe B when he disingenuously avows the gender identity ‘Woman’.

In fact, it is rational to doubt B’s avowals on the account given. Because B’s avowals are obviously intended as part of a political stunt targeting trans people, one is justified in forming the belief that those avowals are not sincere. This does not require us to first come to a rationally justified belief about B’s gender identity, as it did in the case of A. Instead, warrant for the belief that B is insincere comes from knowledge of B’s intentions, as well as B’s views regarding trans people, which are obtained independently of beliefs about B’s gender identity.

The epistemological waters here have the potential to get very murky. For example, if B later comes to realise through introspection that they are, in fact, transgender, at what point does it become irrational to disbelieve their avowals? Pending a fuller analysis of the epistemology of gender avowals, the non-representationalist account is unlikely, as yet, to yield clear-cut results in these sorts of complex cases. My intention in this section is to show that the account is robust enough to deal with pressing problem cases, which sets it above GID-style views while still guaranteeing FPA in paradigm cases, and sets it on a footing to be expanded upon.

5.3 Competence

Another problem case concerns individuals who lack competence in the gender language game sufficient to avow a gender. Elizabeth Barnes raises the case of severely cognitively disabled people, who, due to disability, are unable either to form or to express a gender identity (Barnes, 2022). Barnes uses this case to argue against views that make gender dependent upon gender identity, first, on the basis that it is dehumanising to say that these people are cognitively incapable of having a gender, and second, because doing so prevents us from understanding the specific gendered injustices perpetrated against these people - primarily cognitively disabled women - and risks deepening those injustices (*ibid.*, §§4-5). Barnes takes this to speak both against theories that make gender identity to be determined by avowals (e.g.: Bettcher, 2009), which cognitively disabled people might not be able to articulate, and against theories that take gender identity to be determined by a cognitively complex relation to gender norms (e.g.: McKitrick, 2015, Jenkins, 2016), since there exist cognitively disabled people who are unable to relate to norms in the relevant ways, on account of their disability. For Barnes, any theory that holds womanhood to be determined by having a particular gender identity is inadequate, because there exist cognitively disabled people who do not have such an identity, but who we ought to count as women.

Barnes' paper represents a potent challenge to identity-based accounts of gender generally, and deserves more discussion than I can give it here. However, I will attempt to make clear some points of agreement between her argument and the non-representationalist theory I have presented, and to push back on some of the conflicting points.[7]

Barnes' point about understanding the injustices faced by cognitively disabled people is well taken. As Barnes points out, cognitively-disabled people are far more likely to suffer specifically gendered oppression than non-disabled people. This particularly applies to oppression that construes and targets these people *as women*, such as enforced termination of pregnancies and sterilisation, and rape (Barnes, 2022, pp13-16). I accept completely Barnes' argument that we can't fully conceptualise the injustices suffered by cognitively disabled people unless we regard them specifically as *gendered* injustices.

The point of difference is the further argument that, because these injustices are gendered, we ought to regard the victims as being correspondingly gendered. On the face of it, this looks problematic; a trans man, for instance, could suffer any of the gendered forms of oppression Barnes lists (and many have), without this implying that he is a woman. The most obvious solution is to apply Barnes' own concepts "masculinized" and "feminized", from an earlier paper (Barnes, 2019, pp11-12). These are defined in terms of social positioning; one is feminized, for instance, when one is taken to be biologically female, and allotted a social position on this basis, along with the specific expectations and norms that go along with that social position. Crucially, though, being feminized does not imply that one is a woman (*ibid*.). With these concepts, as is the intention when Barnes introduces them, it appears as though we can construe gendered oppression *as* specifically gendered, without thereby assigning genders to the victims.

Regarding the case of cognitive disability, Barnes appears to pre-empt this response, or something like it. Barnes argues that construing gendered oppression in terms of allocated social role and the corresponding social expectations is inadequate to understand the oppression of cognitively disabled women, because such women are often not subject to these social expectations, on account of their disability (Barnes, 2022, p15). Barnes gives the example of a UK court case in which a judge ordered the termination of a pregnancy of a cognitively disabled woman, specifically ruling that, because of her disability, "she was not capable, despite her wishes, of being a *mother*" (*ibid*., emphasis in original).

[7] My sincere thanks to an anonymous reviewer for pushing me to engage more fully with Barnes' arguments.

However, this point is in tension with Barnes' argument that the oppression of these women should be construed as being of a piece with the wider oppression of women generally, often with regards to limiting or removing control over reproduction:

> The person who is told she cannot carry a pregnancy to term because of her cognitive disability is experiencing our entrenched social norms about women and motherhood just as much as the person who is told she's being selfish if she doesn't have children or the person who is told she's being neglectful if she doesn't breastfeed. The person who undergoes a hysterectomy without even being told about the procedure is experiencing the way we remove women from choices about their own bodies just as much as the person who is told she can't have birth control because it will encourage her to sleep around.
>
> Barnes 2022, p15

Barnes wants to situate the gendered oppression of cognitively disabled women within the wider oppression of women in general (an effort I agree with), but going by her description, this seems to be inextricable from construing this oppression as these women's being subjected to societal expectations about womanhood. The person who is ordered to terminate a pregnancy because she cannot be a mother to the child seems to be being subjected to precisely the gendered expectations applied to mothers ("experiencing our entrenched social norms about women and motherhood").

Thus it appears that an account in terms of feminization is apt; if a cognitively disabled person is oppressed because they are treated as a woman, with the according social expectations, we can analyse that injustice in terms of their being *treated as* a woman, without thereby having to attribute a particular gender to them. Conversely, if these social expectations are not a factor, because the person concerned has a cognitive disability so is not subjected to them, it is hard to see how it could be specifically a gendered form of oppression. It appears to be applying gendered expectations - in the relevant sense of applying gendered standards or norms - that makes these forms of oppression gendered. If we can fully conceptualise the gendered oppression of cognitively disabled people without necessarily attributing genders to them, this removes one reason to think that gender cannot be fully determined by gender identity.

The other strand of Barnes' argument regards the troubling potential for dehumanising cognitively disabled people by denying them genders. Barnes notes that there is nothing inherently dehumanising about holding someone to be agender. The important fact here is that certain cognitively disabled people do not simply lack a categorisable gender identity, but are cognitively incapable of having such. Barnes argues that saying that such people do not have a gender is therefore less like calling a non-disabled person agender, and more like saying that an inanimate object or a non-human animal does not have a gender - that is to say, it is dehumanising. It therefore risks othering these people and deepening the injustices they face (*ibid.* pp10-22). Thus we ought to decouple gender from gender identity on ameliorative grounds.

Since this argument relies on analogy, it is difficult to approach directly. To a degree the non-representationalist has to bite the bullet: Certain cognitively disabled people are incapable of interacting with gendered norms and practices, and are therefore incapable of forming a gender identity, and thus of having a gender, on the current theory. As Barnes points out, dehumanising comparisons can be drawn on this basis, but this is not to say that these comparisons are apt. On identity-based views of gender, to say that someone is gender G is just to say that they have gender identity ‘G’. Thus to say that a severely cognitively disabled person is incapable of having a gender is to say no more than that they are unable to interact with gendered behavioural norms in the relevant way, which is granted by all parties.

All the identity theorist of gender can do is object to the comparisons being drawn. *In this specific respect*, a severely cognitively disabled person might be alike to a computer, to use Barnes’ example, but they differ in many more important ways - most obviously, in that they are a human being, and worthy of respect on that basis. It is the comparison itself that is dehumanising, in that it invites one to associate unrelated non-human traits with the trait of being unable to have a gender. We ought vehemently to reject these associations, but this does not itself require rejecting an identity-based theory of gender.

Note that, in this instance, we are not talking about people who are incapable simply of verbally avowing a gender, but, as Barnes puts it, people for whom there is no “secret ingredient” or hidden set of facts that underpins a gender identity (*ibid.* p21). This is an important distinction. Normally, an individual’s first personal authority ought rationally to override the judgements of others if the individual is able to make a sincere avowal, as per Inf(1) and Inf(2). However, if no such avowals can be obtained, for instance, because they are unable to verbalise such an avowal because of cognitive disability, there is nothing preventing an observer from forming their own rationally-supported judgements about that individual’s gender identity.[8]

Since gender identity is a para-social property, and per the discussion of felt-relevancy and fittingness above, the observer should take into account whether the individual takes it to be appropriate that they be judged by certain gender norms and not others, and use this to inform their judgement. As Barnes notes, caregivers and close acquaintances are often adamant that cognitively disabled people have genders (*ibid.*, p22). On the non-representationalist account, these beliefs can certainly be rationally justified on this basis. However, in cases where a cognitively disabled person is unable to understand or interact with gender norms, there does not seem to be a warranted basis for attributing a particular gender to them.[9]

[8] On the account as given, it is *only* in cases where no (sincere) avowals are obtained that an observer’s judgements about an individual’s relation to gender norms are sufficient to justify beliefs about their gender identity. This is due to Inf(2), which establishes, effectively, that a subject’s sincere avowals that they are gender G override any justification for any contrary beliefs, thus maintaining FPA (see also Sect. 5.1 above). Thus it is only in the absence of sincere avowals that other forms of justification become relevant.

[9] Note that Barnes attributes the push by close acquaintances to attribute gender to cognitively disabled people partly to the desire to include them in feminist discourse, which has historically erased them (Barnes, 2022, p22). If the above argument is granted, then this can be achieved by paying attention to how these people are oppressed on the basis of being feminized.

6 Objections

In this section, I will consider objections to the non-representationalist theory I have put forward, and attempt to respond to them in an informative way.

6.1 Playing along and anti-realism

First I feel I must comment to forestall an objection to the account that arises from a potential misconception. Because the non-representationalist take on gender vocabulary offers no complete metaphysical analysis of gender, and instead offers an analysis of gender terms, the rules and permissions attached to them, etc., it is tempting to construe it as an account merely of what one has to do in order to "play along", as Gus Turyn (2023) puts it, with acknowledging someone's gender identity, rather than as an account of anything deeper. It would be a mistake to treat the account as such. This is most simply shown by noting that the inferential conditions above are intended not merely to determine what it is permissible to say, but to determine warrant or justification, in the sense required for knowledge. The inferential conditions on inferences relating to gender terms do constitute permissions and obligations, but these are the permissions and obligations we are subject to *qua* reasoners. The playing along problem is that someone could continue to use gender vocabulary in an accommodating way in respect of S, while still holding private beliefs that S is not the gender they claim to be (*ibid.*). On the non-representationalist account I have given, such private beliefs are epistemically unwarranted. This is as far as we can go in honouring FPA; we cannot provide an account on which such beliefs cannot be held, only one on which one fails to act as a minimally good reasoner if one holds them.

In a similar vein, it is often thought that deflationist accounts such as the one offered here have something of anti-realism about them. For instance, Bar-On and Long (2001) attribute such to the deflationist approach to first-personal avowals of mental states, on which the present view is based: Quoting Crispin Wright, the authors argue that the deflationary view denies that first-person avowals represent "cognitive achievements", and that they are therefore "cognitively insubstantial", and that mental ascriptions generally (including second- and third-person ones) are "not accountable to any reality." (Wright, 1986, quoted in Bar-On & Long, 2001, p320) Similarly, Elizabeth Barnes objects that a deflationary approach is unable to account for the character of the debate surrounding gender terms, e.g.: the fact that biological essentialism about gender fails not because it has undesirable consequences, but because it fails to describe reality (Barnes, 2017, p2420).

These objections have in common a perception of metaphysical deflationism as trying to unmoor language from reality, such that sentences do not depend on reality for their truth. However, this is a misperception. As Esa Díaz-León points out, in reply to Barnes, deflationism relies on two approaches to answer metaphysical questions: Conceptual analysis and empirical investigation (Díaz-León, 2018, p204). Conceptual analysis investigates and clarifies what we mean by our terms, while empirical investigation applies those terms, with those meanings, to investigate the world, thus engaging with reality in the operative sense (*ibid.* p213; *cf.* Thomasson, 2016, 2020). We unavoidably have to make a choice about what inference rules our terms should follow, but once we do, sentences such as “Trans women are women” are true because they correctly chart a structure of reality.

6.2 Deflationism vs fictionalism

The key weakness of the deflationist account of gender terms appears to be that it makes a metaphysical explanation of gender properties impossible, since it throws such explanation back on gender identity, which does not admit of a substantive, complete metaphysical definition. It may thus appear that the account is theoretically lacking, since it doesn’t actually explain anything - it is in fact giving up on finding an explanation. I will briefly make the case that this is untrue, by comparing the deflationist account to gender fictionalism.

Gender fictionalism accepts that a metaphysical account of gender properties is impossible because, ultimately, the problem of articulating what properties gender terms represent cannot be solved. Nevertheless, it aims to provide an account of the usage of our gender terms by cashing out their rules of use as the rules of a fiction-creating practice. Hence the gender fictionalist is able to do explanatory work regarding gender terms, even though they think there is no coherent metaphysical explanation to be given regarding gender properties, because they can analyse the rules of gender-related linguistic practices, as contributing to the gender fiction. Further, they can look beyond the confines of those rules to how adherence to the gender fiction affects people and their place in society. There is, in fact, a wealth of explanation to be had regarding gender, even if we accept, like the fictionalist, that no substantive metaphysical explanation can be given.

Deflationism about gender enjoys these same explanatory benefits. The difference between it and fictionalism is that the deflationist rejects the principle that if a metaphysical explanation of a property cannot be given, then the property must not really exist. That is, the deflationist accepts the existence of metaphysically inscrutable properties. Hence the deflationist can accept the fictionalist model, but views the rules of use for gender terms as the rules of a language game, not of a fiction. The same examination of the social role of gender can be undertaken, and in the same way (indeed, the EMU reserves a clause for exactly this), but with the added benefit that attributions of gender become truth-apt.

This benefit is considerable. The fact that (cis and trans) individuals’ self-ascriptions of gender are assessable for truth means that our theorising about the rules of use for gender terms is theorising about the conditions of justification *and truth conditions* of beliefs about gender, rather than outlining the rules of a fiction. As noted, Kapusta takes this factual recognition to be a pressing goal for a theory of gender (Eckstrand & Kapusta, 2018).

Logue attempts to minimise the cost of the anti-realism of gender fictionalism by emphasising that the view makes no distinction between trans people and cis people; everyone’s gender ascriptions are to be taken as fictional. However, the issue makes a difference to the aims of an ameliorative account of the metaphysics of gender. Fictionalising the property of gender affects trans people disproportionately to cis people by pathologising their pursuit of proper gender recognition. Logue is aware of the problem, and responds that gender properties being fictional doesn’t mean that peoples' choices with regards to their gender expression ought not to be respected (Logue, 2022, pp157-9). However, Logue's argument does not fully capture the relevant asymmetry. Trans people frequently take on significant risks and burdens, often uprooting their lives in pursuit of proper recognition of their gender, while cis people typically do not. If we adopt a metaphysics on which gender is merely fictional, then we make trans people out to be exposing themselves to harms in pursuit of a property that doesn't exist, in contrast to cis people, who do not generally do so voluntarily.

It follows that attributing genuine metaphysical significance to gender claims is a highly desirable feature of an account of gender vocabulary and gender properties. If such an account of gender self-ascriptions were impossible, then we might abandon the requirement, and adopt a gender fictionalism. However, such an account is only impossible if we unquestioningly adopt the representationalist assumption. Since we can reject the representationalist assumption, we have an account that can guarantee the metaphysical reality of gender, and the metaphysical weight of gender ascriptions.

7 Conclusion

I have argued that there exists a plausible position that is passed over by the current literature on the metaphysics of gender. That position - a deflationism about gender properties - comes bound up with a deflationary metasemantic account of gender vocabulary, namely non-representationalism. I have laid out how the account works, showing how the non-representationalist constructs a language game for gender vocabulary that guarantees an epistemic norm of first-person authority (by ensuring that, in ordinary cases, an individual's avowal of their own gender ought rationally to be respected), and how it solves the inclusion problem (by ensuring that, for any gender identity 'G', everyone who has that gender identity is correctly classed as a G). To my knowledge, the non-representationalist stance is the only one able to satisfy these two competing needs.

The theoretical cost of the stance is obvious: One has to maintain that gender is, to a degree, metaphysically inscrutable. However, this cost can be mitigated. There are illuminating metaphysical things we can say about gender identity, upon which gender depends. For instance, it is still illuminating to think of gender identity as a para-social relational property, i.e.: neither wholly a social classification dependent entirely upon facts about social positioning, nor an entirely subjective (or personal) property dependent only upon facts about the self, but rather a relation between these two types of features. Fleshing out gender identity as a para-social relation would be a fruitful avenue for future work within the framework I have outlined here.

The theoretical cost also seems to be deserved. Given the state of the literature on the metaphysics of gender, and the inclusion problem specifically, gender appears to be a metaphysically inscrutable property *par excellence* - at least in the sense outlined here, namely that it doesn't admit of a representationalist analysis. If we allow the possibility of such properties, which admit only of deflationist explication, then there appears no reason not to grant that gender is like this. As Logue's and Bogardus' arguments show, there is simply no other way that gender properties could do everything we need of them. I have thus argued that a non-representationalism about gender vocabulary is not only a plausible stance, but a highly desirable one. My hope is that the availability of such a position can help to stabilise the discussion on the metaphysics of gender, and open up new avenues for illuminating facts about gender properties, construed in a deflationary way.

Acknowledgements I am grateful to the Humanities Research Centre at the University of York for financial assistance, and to colleagues at the University of York for many helpful discussions. I would also like to thank two anonymous referees at *Synthese* for their insightful comments.

Declarations

Conflict of interest The author has no relevant financial or non-financial interests to disclose in relation to the content of the paper.

References

Antony, L. (2020). Feminism without metaphysics or a deflationary account of gender. *Erkenntnis,85*(3), 529–549. https://doi.org/10.1007/s10670-020-00243-2

Ásta, Á. (2018). *Categories we live by: The construction of sex, gender, race, and other social categories*. Oxford University Press.

Barnes, E. (2017). Realism and social structure. *Philosophical Studies,174*(10), 2417–2433. https://doi.org/10.1007/s11098-016-0743-y

Barnes, E. (2019). Gender and gender terms. *Noûs,54*(3), 704–730. https://doi.org/10.1111/nous.12279

Barnes, E. (2022). Gender without gender identity: The case of cognitive disability. *Mind,131*(523), 838–864. https://doi.org/10.1093/mind/fzab086

Bar-On,D.(2004).*Speakingmymind*.OxfordUniversityPress.https://doi.org/10.1093/0199276285.003.0010

Bar-On, D., & Long, D. C. (2001). Avowals and first-person privilege. *Philosophy and Phenomenological Research,62*(2), 311–335. https://doi.org/10.2307/2653701

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

Bogardus, T. (2019). Some internal problems with revisionary gender concepts. *Philosophia,48*(1), 55–75. https://doi.org/10.1007/s11406-019-00107-2

Bogardus, T. (2022). Why the trans inclusion problem cannot be solved. *Philosophia,50*(4), 1639–1664. https://doi.org/10.1007/s11406-022-00525-9

Boghossian, P. A. (1990). The status of content. *The Philosophical Review,99*(2), 157. https://doi.org/10.2307/2185488

Boghossian, P. A. (2002). The rule-following considerations. In A. Miller (Ed.), *Rule-following and meaning* (pp. 141–187). Acumen Publishing.

Brandom, R. (1994). *Making it explicit*. Harvard University Press.

Carnap, R. (1950). Empiricism, semantics and ontology. *Revue Internationale De Philosophie,4*(11), 20–40.

Dembroff, R. (2018). Real talk on the metaphysics of gender. *Philosophical Topics,46*(2), 21–50. https://doi.org/10.5840/philtopics201846212

Díaz-León, E. (2018). On Haslanger's meta-metaphysics: Social structures and metaphysical deflationism. *Disputatio,10*(50), 201–216. https://doi.org/10.2478/disp-2018-0013

Díaz-León, E. (2020). Descriptive vs. ameliorative projects: The role of normative considerations. In A. Burgess, H. Cappelen, & D. Plunkett (Eds.), *Conceptual engineering and conceptual ethics* (pp. 170–186). Oxford University Press.

Eckstrand, N., & Kapusta, S. (2018). Trans*feminism: How trans* issues and feminism overlap. *Blog of the APA*. Available at: https://blog.apaonline.org/2018/08/09/transfeminism-how-trans-issues-and-feminism-overlap/. Accessed 14 Mar 2024.

Fan, N. (2022). Why Avowals must be assertions. *Philosophical Investigations,46*(2), 221–239. https://doi.org/10.1111/phin.12369

Fricker, M., & Jenkins, K. (2017). Epistemic injustice, ignorance, and trans experiences. In A. Garry, S. J. Khader, & A. Stone (Eds.), *The Routledge companion to feminist philosophy* (pp. 268–278). Routledge.

Haslanger, S. (2000). Gender and race: (What) are they? (What) do we want them to be? *Noûs,34*(1), 31–55. https://doi.org/10.1111/0029-4624.00201

Haslanger, S. A. (2012). *Resisting reality: Social construction and social critique*. Oxford University Press.

Haslanger, S. (2017). The sex/gender distinction and the social construction of reality. In A. Garry, S. J. Khader, & A. Stone (Eds.), *The Routledge companion to feminist philosophy* (pp. 157–167). Routledge.

Jenkins, K. (2016). Amelioration and inclusion: Gender identity and the concept of woman. *Ethics,126*(2), 394–421. https://doi.org/10.1086/683535

Jenkins, K. (2018). Toward an account of gender identity. *Ergo an Open Access Journal of Philosophy*, *5*(20201214). https://doi.org/10.3998/ergo.12405314.0005.027

Kapusta, S. (2016). Misgendering and its moral contestability. *Hypatia, 31*(3), 502–519. https://doi.org/10.1111/hypa.12259

Kirkland, K. L. (2018). Feminist aims and a trans-inclusive definition of "woman". *Feminist Philosophy Quarterly, 5*(1). https://doi.org/10.5206/fpq/2019.1.7313

Logue, H. (2022). Gender fictionalism. *Ergo an Open Access Journal of Philosophy*, *8*(28). https://doi.org/10.3998/ergo.2229

Lugones, M. (2003). *Pilgrimages/peregrinajes: Theorizing coalition against multiple oppressions*. Rowman & Littlefield.

McKitrick, J. (2015). A dispositional account of gender. *Philosophical Studies,172*(10), 2575–2589. https://doi.org/10.1007/s11098-014-0425-6

Mikkola, M. (2009). Gender concepts and intuitions. *Canadian Journal of Philosophy,39*(4), 559–583. https://doi.org/10.1353/cjp.0.0060

Mikkola, M. (2016). *The wrong of injustice: Dehumanization and its role in feminist philosophy*. Oxford University Press.

Price, H. (2011). *Naturalism without mirrors*. Oxford University Press.

Price, H. (2013). *Expressivism*. Cambridge University Press.

Ritchie, K. (2021). Essentializing inferences. *Mind & Language,36*(4), 570–591. https://doi.org/10.1111/mila.12360

Rowland, R. A. (2023a). Recent work on gender identity and gender. *Analysis,83*(4), 801–820. https://doi.org/10.1093/analys/anad027

Rowland, R. A. (2023b). The normativity of gender. *Noûs,58*(1), 244–270. https://doi.org/10.1111/nous.12453

Simpson, M. (2019). What is global expressivism? *The Philosophical Quarterly,70*(278), 140–161. https://doi.org/10.1093/pq/pqz033

Simpson, M. (2020). Creeping minimalism and subject matter. *Canadian Journal of Philosophy,50*(6), 750–766. https://doi.org/10.1017/can.2020.20

Thomasson, A. L. (2016). Metaphysical disputes and metalinguistic negotiation. *Analytic Philosophy,58*(1), 1–28. https://doi.org/10.1111/phib.12087

Thomasson, A. L. (2020). A pragmatic method for normative conceptual work. In A. Burgess, H. Cappelen, & D. Plunkett (Eds.), *Conceptual engineering and conceptual ethics* (pp. 435–458). Oxford University Press.

Tiefensee, C. (2018). Saving which differences? Creeping minimalism and disagreement. *Philosophical Studies,176*(7), 1905–1921. https://doi.org/10.1007/s11098-018-1103-x

Turyn, G. (2023). Gender and first-person authority. *Synthese,201*(4), 1–19. https://doi.org/10.1007/s11229-023-04125-2

Williams, M. (2013). How pragmatists can be local expressivists. *Price 2013, expressivism, pragmatism and representationalism* (pp. 128–144). Cambridge University Press.

Williams, M. (2015). Knowledge in practice. In D. K. Henderson & J. Greco (Eds.), *Epistemic evaluation: Purposeful epistemology* (pp. 161–185). Oxford University Press.

Wittgenstein, L. (1950). *Philosophical Investigations*. Blackwell.

Wright, C. (1986) On making up one's mind: Wittgenstein on intention. In Weingartner, P., & Schurz, G. (Eds.) *Logic, Philosophy of Science and Epistemology, Proceedings of the 11th International Wittgenstein Symposium, Kirchberg, Vienna*.

Zalabardo, J. L. (2019). The primacy of practice. *Royal Institute of Philosophy Supplement,86*, 181–199. https://doi.org/10.1017/s1358246119000122