Gender First: A Theory of Binary Category Terms

Cameron Domenico Kirk-Giannini

[Draft as of 4/30/2023]

Abstract: Let the label *binary category terms* refer to natural language expressions like 'woman', 'man', 'female', and 'male'. Focusing on 'woman' and 'female', I develop a novel, empirically supported theory of the meanings of English binary category terms. Given plausible assumptions about the metaphysics of sex and gender, this *gender-first* theory predicts that the sentence 'Trans women are women' expresses a truth in all contexts and the sentence 'Women are adult human females' expresses a truth in most ordinary contexts — thus that these two sentences can and usually do express logically consistent contents. The key feature of the proposed theory is that it treats both 'woman' and 'female' as sensitive to an individual's gender when that individual belongs to a gendered category and to an individual's sex otherwise. The existence and plausibility of a gender-first theory of this kind opens up conceptual room for trans-inclusive positions in the philosophy of sex and gender which endorse the claim that women are adult human females, thereby both accounting for trans women's experiences of their bodies as female and helping to disarm the sentence 'Women are adult human females' as a trans-exclusionary slogan.

Feminist philosophical work on sex and gender has typically held that the claim that trans identities are valid is not consistent with the claim that women are adult human females and men are adult human males. If trans identities are valid, the reasoning supporting this conclusion goes, then trans women are women and trans men are men. But trans women need not be female and trans men need not be male. Thus, to claim that women are adult human females is to exclude some or all trans women — those who are not female — from the class of women. The same is true, mutatis mutandis, for the claim that men are adult human males. Because of this inconsistency, the claim that women are adult human females has come to be regarded as inherently exclusionary.[1]

The conclusion that we cannot both affirm trans identities and hold that women are adult human females and men are adult human males — to which I will refer simply as INCONSISTENCY — is unwelcome for three reasons.[2] First, there is linguistic evidence suggesting that, at least for many English speakers, both the sentence 'Trans women are women' and the sentence 'Women are adult human females' express truths in most contexts. Second, trans individuals often describe themselves

---

[1] Here and in what follows, bare plurals should be treated as implicitly universally quantified. Thus, when I write that *trans identities are valid*, I should be understood as meaning that *all* trans identities are valid, when I write that *trans women are women*, I should be understood as meaning that *all* trans women are women, and so on.

[2] From a logical point of view, the tension between affirming trans identities and holding that trans women are women and trans men are men only arises in the presence of the assumption that there are trans women and men — otherwise the various universally quantified claims could all be vacuously true at the same time. Since this assumption, which we might call EXISTENCE, is uncontroversial, I will leave it implicit in what follows.

and their bodies as female (in the case of trans women) or male (in the case of trans men). The most straightforward way to explain this practice is to hold that trans women are female and trans men are male. Third, because the sentence 'Women are adult human females' strikes many ordinary speakers of English as obviously true and even platitudinous, endorsing INCONSISTENCY opens up rhetorical space for using its plausibility to depict trans-inclusive projects as confused, unintuitive, or obviously mistaken.

Happily, I argue in what follows, the empirical case for INCONSISTENCY is weak. Focusing for simplicity on women, and setting aside for the time being the complexities introduced by context sensitivity, there is no tension between the claim that trans women are women and the claim that women are adult human females when 'female' and 'woman' are given their ordinary interpretations in the idiolects of many English speakers.[3,4] On the contrary, it is possible to offer an empirically supported theory of meaning for 'woman,' 'female,' and related expressions according to which, given certain plausible assumptions about the metaphysics of sex and gender, both 'Trans women are women' and 'Women are adult human females' express truths. The key to constructing such a theory is to hold that both 'woman' and 'female' express complex properties which pick out an individual's gender if that individual belongs to a gendered category and an individual's sex otherwise. Adopting a *gender-first* semantics of this kind both makes sense of trans individuals' experiences of themselves and helps to neutralize the sentence 'Women are adult human females' as an exclusionary slogan: Since one need not deny the validity of trans identities to accept it as true, whatever intuitive plausibility it enjoys fails to constitute evidence for an exclusionary metaphysics. And it does this without positing radical context sensitivity or ambiguity in the meanings of 'female' or 'woman'. The gender-first approach to the meanings of 'female' and 'woman' thus constitutes a new theoretical option with advantages over existing alternatives.

The availability of a gender-first theory of the meanings of expressions like 'woman' and 'female' means that it would be inapt to refer to them either as *sex terms* or as *gender terms*. Instead, I will refer to the expressions 'woman', 'man', 'female', 'male', 'girl', and 'boy' neutrally as *binary category terms*. My choice of this terminology reflects the descriptive fact that in natural language these terms usually come in pairs.[5]

My discussion is organized as follows. I begin by describing in more detail the kind of reasoning which has led philosophers of sex and gender to endorse INCONSISTENCY and situating my project within the broader feminist literature on these topics (Section 1). I then outline the empirical

---

[3] My focus on 'woman' and 'female' can be justified in two ways. First, the dialectic surrounding 'woman' and 'female' can be reformulated, mutatis mutandis, for 'man' and 'male', and this change does not substantially affect the force of my arguments. Second, because they are more directly implicated in feminist issues, 'woman' and 'female' have received considerably more attention in the recent literature than 'man' and 'male'.

[4] An English-speaking individual's *idiolect* is their own personal dialect of the English language.

[5] I do not mean to suggest, for example, that every adult does or ought to belong either in the extension of 'woman' or in the extension of 'man'. Indeed, I think the claim that binary category terms neatly divide humans into mutually exclusive and exhaustive categories must be rejected in light of the existence of intersex and non-binary individuals. Nevertheless, the terms themselves are binary in the sense I intend.

motivations for holding that the sentences 'Trans women are women' and 'Women are adult human females' express truths in the idiolects of many English speakers (Section 2). After discussing two further motivations for rejecting INCONSISTENCY, Section 3 turns to semantics. There, I introduce a range of simple candidate meanings for 'woman' and 'female' and assess the viability of three well-known semantic theories which can be constructed by combining them. Finding none of these theories empirically adequate, I present my gender-first alternative in Section 4. Sections 5 and 6 discuss the evidence that 'female' is ambiguous, polysemous, or context sensitive and consider its implications for the argument of the previous sections. Section 7 defends my theory against potential objections. Section 8 concludes by summarizing the advantages of the gender-first approach as compared to existing theories.

1. The Default Position and the Sex-Only View

I have sketched the reasoning which has led many philosophers of sex and gender to accept INCONSISTENCY. In this section, I add detail to the picture by describing two contrasting packages of views about sex, gender, and language which have been prominent in recent philosophical discussions and which entail INCONSISTENCY.

The first package of views has come to be something like the default position in philosophical discussions of sex and gender — indeed, I will refer to it in what follows as the *Default Position*.[6] According to the Default Position, sex and gender are metaphysically distinct. An individual's sex is one of their biological properties (akin to their eye color or blood type), whereas an individual's gender is one of their social properties (akin to their socioeconomic status or profession). Because there is no necessary connection between an individual's biological properties and their social properties, individuals of a given sex can vary with respect to gender and individuals of a given gender can vary with respect to sex.

Moreover, the Default Position holds that the metaphysical distinction between sex and gender is mirrored by a linguistic one: 'female' and 'male' are sex terms, while 'woman' and 'man' are gender terms. For clarity, let us adopt a convention of referring to the biological sex that characteristically produces large, immobile gametes as *Sex F* and to individuals of this sex as *F-sexed*, while correspondingly referring to the biological sex that characteristically produces small, mobile gametes as *Sex M* and to individuals of this sex as *M-sexed*. Then we can understand the Default Position as holding that the English expressions 'female' and 'male' pick out the properties *being F-sexed* and *being M-sexed*, respectively.

Similarly, if we refer to the gender typically associated with F-sexed individuals as *Gender W* and to individuals possessing this gender as *W-gendered*, and if we refer to the gender typically associated with M-sexed individuals as *Gender M* and to individuals possessing this gender as *M-gendered*, we can understand the Default Position as holding that the English expressions 'woman' and 'man' pick out

---

[6] Throughout, I adopt the typographical convention of putting the names of significant theoretical propositions in small caps (e.g. INCONSISTENCY), the names of theories themselves in title case (e.g. the Default Position), and the names of candidate lexical entries for particular words in parentheses (e.g. ('Woman': Gender Only) below).

the properties *being a W-gendered adult* and *being an M-gendered adult* respectively.[7] It thus follows from the Default Position that the English sentences 'Not all women need be females' and 'Not all males need be men' express truths.[8]

There is much to recommend the Default Position. It offers a natural explanation of what the claims of trans women to be women mean: When trans women claim that they are women, they are making a claim about their social properties rather than their biological properties. Combined with the plausible metaphysical thesis that trans women do possess the relevant social properties, the Default Position also explains why their claims to be women are correct. Moreover, the Default Position explains how it is possible for an individual to be non-binary with respect to gender but not sex: Such a person might possess a sex but fail to possess the social property corresponding to any gender. And it explains how it is possible for an individual to be intersex but be a man or a woman: Such a person might fail determinately to belong to any sex but nonetheless possess the social property corresponding to a gender.[9]

---

[7] While introducing new terminology for referring to sexes and genders is somewhat cumbersome, it is superior in this context to the main alternative, which is to use 'female' and 'male' stipulatively to pick out sex categories and 'woman' and 'man' stipulatively to pick out gender categories. Talking in this alternative way can misleadingly make certain substantive semantic theses seem tautological. For example, without introducing new terminology, the thesis that the English expression 'female' picks out a sex category can only be stated using apparently trivial sentences like *'female' picks out the property of being female.*

[8] A clear recent statement of the Default Position can be found in Mikkola (2022):

> "Historically many feminists have understood 'woman'… not as a sex term, but as a gender term that depends on social and cultural factors (like social position). In so doing, they distinguished sex (being female or male) from gender (being a woman or a man)…"

For other statements, see, for example, Stoljar (1995), Haslanger (2000), and Dembroff (2016); Rubin (1975) also contains a statement of the position, though she uses the terms 'sex' and 'gender' somewhat flexibly. Note that my claim that this package of views is the default position in philosophical discussions of sex and gender is not meant to imply that it is endorsed by all or even most theorists working in this area. There are those who reject the idea of sex as a biological property (e.g. Butler (1990, 1993)), those who reject the idea of gender as a social property independent of sex (e.g. Lawford-Smith (2022)), those who defend context-sensitive accounts of terms like 'female' and 'woman' on which they do not work in the straightforward way described by the Default Position (e.g. Barnes (2020)), and those who reject the project of giving a semantics for binary category terms altogether (e.g. Mikkola (2016)).

[9] Though I do not endorse the Default Position, I share its basic commitment to the existence of at least two sexes and at least two genders. While there is considerable controversy regarding the correct metaphysical accounts of sex (see e.g. Ásta (2011), Butler (1990, 1993), Dembroff (2021), and Fausto-Sterling (1993, 2000)) and gender (see e.g. Ásta (2011, 2013, 2018), Bach (2012), Barnes (2020, 2022), Butler (1990), Haslanger (2000), Jenkins (2016), McKitrick (2015), Stoljar (1995, 2011), and Witt (2011a, 2011b)), the position that there are sexes and genders is less contentious.

I also assume that some trans women are not F-sexed and some trans men are not M-sexed. I will call this assumption DIVERSITY. While there is substantial disagreement about the correct metaphysical account of sex, there is equally substantial consensus regarding DIVERSITY. (Note that

To see that the Default Position is committed to INCONSISTENCY, note that, given her understanding of the meanings of 'woman' and 'female', the defender of the Default Position who recognizes the validity of trans identities must endorse the sentence 'Some women are not females'. And this sentence is straightforwardly inconsistent with the sentence 'Women are adult human females' when the latter is understood as implicitly universally quantified.[10]

Despite its theoretical advantages, the Default Position has recently become the subject of controversy. On the basis of a number of arguments turning on linguistic and metaphysical intuitions, it has been claimed that the idea that 'woman' and 'man' are gender terms is mistaken. Instead, it has been claimed, they are sex terms, so that what it is to be a woman is to be an F-sexed adult human and what it is to be a man is to be an M-sexed adult human.[11] In what follows, I will refer to the view that sex is a biological property and both 'woman' and 'female' are sex terms as the *Sex-Only View*.

To see that the Sex-Only View is committed to INCONSISTENCY, note that it is committed to the truth of the sentence 'Women are adult human females' because it holds that both 'woman' and 'adult human female' pick out the property of being an F-sexed adult human. Since not all trans women are F-sexed, this commitment to the truth of 'Women are adult human females' is incompatible with the truth of 'Trans women are women'.[12]

There are thus many roads to INCONSISTENCY. Indeed, the idea that there is some deep tension between the claim that trans women are women and the claim that women are adult human females has come to dominate philosophical discussions of sex and gender. Even accounts of the meanings of 'woman' and 'female' which posit pervasive context sensitivity, such as the account defended by Barnes (2020), do not question the assumption that 'Women are adult human females' and 'Trans women are women' express inconsistent propositions in any given context.

When one approaches the philosophy of sex and gender from the perspective of INCONSISTENCY, one is faced with an unpleasant choice: On the one hand, one can follow the Default Position in vindicating the claim that trans women are women but not the claim that women are adult human females; on the other hand, one can follow the Sex-Only View in vindicating the claim that women are adult human females but not the claim that trans women are women.

Given the social and political ramifications of abandoning the claim that trans women are women, most feminist philosophers have chosen the first of these options over the second.[13] For the same

---

DIVERSITY is *not* equivalent to the claim that some trans women are not female or the claim that the sentence 'Trans women are female' expresses a falsehood in English.)

[10] Strictly speaking, my claim is that the Default Position is committed to INCONSISTENCY in the presence of EXISTENCE and DIVERSITY.

[11] See for example Byrne (2020, 2021, 2022).

[12] Again, my claim is that the Sex-Only view is committed to INCONSISTENCY in the presence of EXISTENCE and DIVERSITY.

[13] Though see Byrne (2020, 2021, 2022), Lawford-Smith (2022).

reasons, the Sex-Only View has been criticized as exclusionary, and this criticism has extended to the claim that women are adult human females.

Yet there is a way out of the dilemma. We need not choose between recognizing the validity of trans identities and holding that women are adult human females. If we allow that both 'woman' and 'female' can sometimes pick out gendered categories, we can reject INCONSISTENCY and hold that both 'Women are adult human females' and 'Trans women are women' express truths. In the next three sections, I defend an account of the meanings of binary category terms with this feature and explain its empirical and practical advantages over both the Default Position and the Sex-Only View.

Before proceeding, I should note that I am not the first to propose the idea that 'male' and 'female' are not always sex terms. One finds suggestions of this idea at least as early as Butler (1990).[14] Nor am I the first to appreciate that this possibility opens up room for an account which vindicates both the idea that trans women are woman and the idea that women are adult human females. This theoretical option has recently been suggested, though not developed, by Heartsilver (2021). My contribution in what follows is to explicitly state and linguistically motivate a novel theory of this kind, as well as to defend it against certain recent criticisms due to Byrne (2020, 2022). To my knowledge, no one has undertaken this project before.[15]

2. INCLUSION and DEFINITION

We have reason to reject INCONSISTENCY only if we have reason to endorse both the claim that trans women are women and the claim that women are adult human females. Before turning to specific semantic proposals, then, it will be helpful to gather together some linguistic data about these two claims. This kind of data serves as the primary source of evidence against which we will judge competing proposals about the meanings of binary category terms. I will refer to the claim that

[14] For example, Butler writes, "perhaps this construct called 'sex' is as culturally constructed as gender; indeed, perhaps it was always already gender, with the consequence that the distinction between sex and gender turns out to be no distinction at all" (1990, 9). Though Butler is here primarily concerned with the metaphysics of sex and gender rather than the meanings of binary category terms, and though her aim is to interrogate the distinction between sex and gender rather than to challenge the idea that categories like *male* and *female* are sex categories, her remarks are at least evocative of the thesis that 'male' and 'female' do not function to pick out biological sexes. See Ásta (2018) for further discussion of Butler's work on sex and gender.

[15] Questions about the meanings of binary category terms can be approached either descriptively or normatively. Descriptive approaches seek to characterize the meanings of expressions as they are actually used in human communities, while normative approaches propose idealized meanings for expressions which, if adopted, would contribute to achieving certain practical goals — usually political goals related to the elimination of injustice (see e.g. Haslanger (2000) and Jenkins (2016) for discussion). My project is descriptive rather than normative: the methods of natural language semantics as a field of inquiry are empirical and therefore appropriate only to descriptive inquiry. And in any case — though I will not defend this position here — it seems to me that normative questions about the meanings of 'female' and 'woman' have less contentious answers: they should be assigned meanings which are trans-inclusive. So it seems to me that, in as much as there are substantive debates to be had about whether 'woman' and 'female' are trans-inclusive, they are descriptive debates.

'Trans women are women' expresses a truth in English as INCLUSION and to the claim that 'Women are adult human females' expresses a truth in English as DEFINITION.[16]

Note that the linguistic intuitions about appropriateness and entailment on which I rely in this section are my own. I take it that these intuitions are not idiosyncratic, however; they are shared by at least a significant proportion of contemporary English speakers. The arguments I make will be most directly relevant to readers who share the semantic intuitions I describe. For them, my arguments will constitute evidence that their idiolect is correctly described by a gender-first semantics for binary category terms. But even readers who do not share all of the semantic intuitions I describe should find the position I develop in what follows interesting, for two reasons. First, it is an empirically plausible theory of the meanings of binary category terms for at least many contemporary English speakers. Second, it constitutes a demonstration that it is possible to give a coherent semantics for binary category terms on which INCONSISTENCY is false.[17]

Thus, while I think that the evidence for INCLUSION and DEFINITION can be appreciated by anyone willing to suspend their pre-existing semantic commitments regarding binary category terms, I do not measure the success of my project according to whether my arguments convince committed defenders of either the Default Position or the Sex-Only View. I count my arguments in what follows a success even if all they are taken to show is that those who endorse INCLUSION need not resist the evidence favoring DEFINITION, and vice versa.

What evidence is there for INCLUSION? Since there can be no doubt that 'Women are women' expresses a truth in English, the question of whether 'Trans women are women' expresses a truth in English turns on the linguistic properties of the adjective 'trans'.

Linguists divide adjectives into three main classes — intersectives, subsectives, and non-subsectives — with this last class containing a further important group known as the privatives.[18] Intersective adjectives form the largest class, comprising ordinary adjectives like 'square', 'hungry', and so forth. They are characterized by their inferential properties: from the fact that something can truly be

---

[16] Until I explicitly address issues of context sensitivity in Section 6, I will move freely between metalanguage and object-language claims. Thus I will sometimes gloss INCLUSION as the claim that trans women are women rather than the claim that 'Trans women are women' expresses a truth in English. Neither of these ways of glossing INCLUSION should be understood to presuppose any substantive theory of the meaning of 'woman' in English.

[17] Some readers may worry that questions about the semantics of binary category terms are so caught up with normative and political questions that the empirical methods of semantics are no longer applicable. If speakers' intuitions about the acceptability of sentences are informed by their normative beliefs, for example, then in what sense can the study of their idiolects be empirical? My response to this kind of worry is that empirical methods can be used to study an individual's idiolect regardless of what causally explains why that individual has that idiolect. It does not matter whether an individual internalizes a particular way of using words on the basis of her normative beliefs; as long as she really does have the relevant linguistic judgments about linguistic acceptability (rather than merely reporting that she has them, for example), empirical methods can be used to study her idiolect. Thanks to [Name] for pressing me to clarify this point.

[18] See, for example, Kamp (1975), Kamp and Partee (1995), Partee (2007), and Del Pinal (2015).

described using 'red square', it follows both that it can correctly be described using 'red' and that it can correctly be described using 'square'. Subsective adjectives, including 'good', 'competent', and so forth, permit the second kind of inference but not the first: from the fact that someone can correctly be described using 'good cook', it follows that they can correctly be described using 'cook', but it does not follow that they can correctly be described using 'good'.[19] Non-subsectives are a heterogeneous group, including both adjectives like 'alleged' (from the fact that someone can truly be described using 'alleged criminal', it follows neither that they can truly be described using 'criminal' nor that they cannot truly be so described) and privatives like 'fake' (from the fact that something can truly be described using 'fake fur', it follows that it *cannot* truly be described using 'fur').

Unless the adjective 'trans' is non-subsective, then, 'Trans women are women' expresses a truth in English in just the same way that 'Red squares are squares' expresses a truth in English. And indeed, there seems to be little evidence that 'trans' is anything other than an intersective adjective. In my own idiolect, for example, (2) is logically entailed by (1)

> (1) Arabella is a trans woman.
> (2) Arabella is a woman.

Of course, entailments like those from (1) to (2) will be contested by those who endorse the Sex-Only View. But it is important not to put theory before data: when studying meaning in natural language, we ought in general move from claims about the entailment relations between expressions or sentences to claims about their meanings, not vice versa. We must not reject the inference from (1) to (2) on the basis of a prior theory of the meaning of 'woman'; we must assess its pre-theoretical plausibility and construct a theory of the meaning of 'woman' responsive to what we find.

In any case, there are other sources of evidence for the claim that 'trans' is not a non-subsective adjective. For example, non-subsective adjectives become marked in the context of taxonomic sentences, as in (4) and (5)

> (3) There are two types of integers: even integers and odd integers.
> (4) ? There are two types of guns: real guns and imaginary guns.
> (5) ? There are three types of offenders: one-time offenders, repeat offenders, and alleged offenders.

Intuitively, (4) and (5) are less acceptable than (3) because they purport to be listing the subsets of a set of objects, but non-subsective adjectives, when they modify a noun, pick out a set of objects which is not a subset of the set of objects picked out by that noun. Because they are marked in this way, we are disposed to interpret a speaker who produces (4) or (5) — but not (3) — as engaging in wordplay, making a joke, or being otherwise nonserious. In this respect, (4) and (5) contrast with:

> (6) There are two types of women: cis women and trans women.

---

[19] To appreciate this fact, consider the invalidity of the argument from 'Jack is a good cook' and 'Jack is a dancer' to 'Jack is a good dancer'.

(6) is not marked in the way it would be if 'trans' were a non-subsective adjective. We have no inclination to interpret a speaker who utters (6) as engaging in wordplay, making a joke, or being otherwise nonserious.

So it seems that, at least in many idiolects, there is linguistic evidence favoring the hypothesis that 'trans' is not a non-subsective adjective. I conclude that 'trans' is not a non-subsective adjective in these idiolects; instead, it seems to be an intersective adjective. It follows that an adequate theory of the meanings of 'trans' and 'woman' must accommodate the datum that 'Trans women are women' expresses a truth in these idiolects.[20]

The evidence for DEFINITION has received attention recently because, combined with the assumption that 'female' is a sex term, DEFINITION has formed part of an argument that 'woman' is a sex term (see, for example, Byrne (2020)).

Perhaps the most significant argument for DEFINITION is also the humblest: it simply strikes many ordinary English speakers as obviously correct. Thus Bettcher writes:

> "On the face of it, the definition 'female, adult, human being' *really does seem right*. Indeed, it seems as perfect a definition as one might have ever wanted." (2009, 105; emphasis in original)

The intuitive appeal of DEFINITION likely explains the fact that it captures the definition of 'woman' in many dictionaries: The Oxford English Dictionary defines 'woman' as 'an adult female human being,' while Merriam-Webster defines 'woman' as 'an adult female person'.[21]

Of course, it would be unwise to base the case for DEFINITION entirely on the intuitions of lexicographers. Further evidence can be found in the fact that, in contexts where our attention is restricted to adult humans, 'woman' and 'female' seem to pick out the same group of people.[22] Thus, in contexts where it is safe to assume that graduate students are adult humans, ordinary utterances of sentences (7) and (8) appear to make exactly the same claim about who came to the colloquium:

(7) All of the graduate student women came to the colloquium.
(8) All of the female graduate students came to the colloquium.

---

[20] There are also normative reasons for endorsing INCLUSION having to do with the harms experienced by trans people when they are denied membership in the gender categories with which they identify (see George and Briggs (2019) and Kirk-Giannini and Glanzberg (fc) for discussion). Since my interest is descriptive rather than normative questions about meaning, however, I set such arguments aside in what follows.

[21] The lexicographic situation is evolving rapidly. In 2020, Merriam-Webster added a trans-inclusive definition for 'female': "having a gender identity that is the opposite of male." And in 2022, the Cambridge Dictionary added a similar definition for 'woman': "an adult who lives and identifies as female though they may have been said to have a different sex at birth." See Kaur (2022) for reporting and discussion.

[22] For further discussion, see Byrne (2020) and Heartsilver (2021).

And this is true even when they are used in a context where it is common ground that not all the W-gendered graduate students are F-sexed and not all the F-sexed graduate students are W-gendered.

DEFINITION neatly explains this observation. If women are adult human females, then when we restrict our attention to adult humans, 'woman' and 'female' must have the same extension. It follows that all women have a property (having come to the colloquium, for example) just in case all females have that property.

I conclude that there is a strong prima facie linguistic case for both INCLUSION and DEFINITION, at least when it comes to the idiolects of many English speakers. This is good evidence that both principles are true in the idiolects of these speakers. Let us call idiolects in which both principles are true *consistent idiolects*. The evidence favoring INCLUSION and DEFINITION provides motivation for developing a semantics for consistent idiolects.

3. Provisional Adequacy and the Gender-Only View

Let us call a semantic theory of binary category terms *provisionally adequate* if it validates INCLUSION and DEFINITION. Provisional adequacy is not a guarantee that a semantic theory for consistent idiolects is defensible all-things-considered; rather, it is offered as a minimum threshold which an empirically supported theory of consistent idiolects ought to meet.

Why develop a provisionally adequate theory of binary category terms? In addition to the linguistic evidence supporting INCLUSION and DEFINITION discussed in the previous section, it is worth mentioning two further motivations.

First, trans men often describe themselves and their bodies as male and trans women often describe themselves and their bodies as female. As Zimman (2014, 22) puts it, "Trans speakers question the reasoning that says particular physiological characteristics are inherently gendered, contesting the assumption that having a penis necessarily makes a body male while having a vagina (or lacking a penis) makes a body female." Focusing on trans men, Zimman reports that they

> "…have in dramatically increasing numbers begun to contest the treatment of their bodies as female and have harnessed the linguistic practices already circulating within their communities in order to discursively reconstruct their bodies as entirely and legitimately male." (2014, 20).

A similar point can be made about language use among trans women. Trans activist and YouTube content creator Natalie Wynn, for example, introduces the idea that trans women have female bodies with her characteristic wit:

> "Now you might be wondering: '…How could I view [the penis] as anything other than a male part? It's a fucking penis'. Well, I'm glad you asked. I always jump at the opportunity to enlighten someone about the ecstasies of the feminine penis." (Wynn 2019)

In contrasting 'feminine' with 'male' and describing the bodies of trans women using the former rather than the latter, Wynn rejects the idea that the applicability of terms like 'male' and 'female' is wholly determined by sex: M-sexed bodies can correctly be described as feminine or female.

Both Zimman and Wynn construe the linguistic practices of trans individuals as oppositional to what they take to be ordinary usage, according to which persons and bodies count as male or female in virtue of facts about genital anatomy. Though I believe there is a use of 'male' and 'female' along these lines (see Section 6 for discussion), the linguistic evidence favoring INCLUSION and DEFINITION suggests that Zimman and Wynn are being too concessive: If ordinary usage supports the idea that trans women are women and women are adult human females, then ordinary usage supports the idea that trans women are females. Thus, adopting a provisionally adequate theory of binary category terms allows us to understand the way trans individuals talk about themselves and their bodies as consistent with, rather than oppositional to, ordinary usage.

A second motivation for developing a provisionally adequate theory of binary category terms is that doing so helps to defuse the rhetorical power of the sentence 'Women are adult human females' as a tool for anti-trans activism. Because INCONSISTENCY has been taken for granted by most philosophers working on sex and gender, it has been possible to use the intuitive appeal of 'Women are adult human females' as the basis for an argument that 'Trans women are women' is false. For this reason, Dembroff (2021, 1001), characterizes the sentence 'Women are adult human females' as "a political slogan that is currently being used to undermine civic rights and respect for trans persons."

One way to respond to the rhetorical use to which the sentence 'Women are adult human females' has been put is to preserve INCONSISTENCY and reject DEFINITION. But this strategy has disadvantages. Many ordinary people are driven away from accepting the claim that trans women are women by the thought that it is incompatible with the claim that women are adult human females: The truth of the latter can seem to many people like a Moorean fact. To my mind, then, a better way to undermine the rhetorical effectiveness of 'Women are adult human females' as an anti-trans slogan is to show that it is compatible with the idea that trans women are women by rejecting INCONSISTENCY and developing a provisionally adequate theory of binary category terms. This strategy acknowledges the plausibility of 'Women are adult human females' but defuses it as a point of contention in arguments about the rights of trans people.[23]

There are thus many reasons — both empirical and practical — to develop a provisionally adequate theory of binary category terms. Nevertheless, most widely endorsed proposals about the meanings of binary category terms fail to be provisionally adequate. To see this, let us first introduce some

_____

[23] Two notes about this consideration in favor of developing a provisionally adequate semantics for binary category terms. First, my suggestion is not that any rhetorical or political consequences follow immediately or inevitably from recognizing the possibility of a provisionally adequate semantics. My suggestion is rather that the realization that a provisionally adequate semantics for binary category terms is empirically supported *makes possible* certain new rhetorical and political strategies. Whether these moves are ever made, and whether they succeed in changing the discourse surrounding trans inclusion, depends on further non-linguistic factors. Second, my suggestion is not that developing a semantics for consistent idiolects will convince committed defenders of (e.g.) the Sex-Only View that what *they* have meant by 'Women are adult human females' is trans-inclusive after all. It is rather that the availability of a provisionally adequate semantics might make trans-inclusive positions more attractive to everyday people who feel the pull of both INCLUSION and DEFINITION and lack any pre-existing theoretical commitments about the semantics of binary category terms. Thanks to [Name] and [Name] for helpful discussion of these points.

simple candidate meanings for binary category terms, operating for the time being under the simplifying assumption that they are not ambiguous, polysemous, or context sensitive.

> ('Woman': Gender Only): 'woman' picks out the property of being a W-gendered adult.
> ('Woman': Sex Only): 'woman' picks out the property of being an adult human of Sex F.

> ('Female': Gender Only): 'female' picks out the property of being W-gendered.
> ('Female': Sex Only): 'female' picks out the property of being F-sexed.[24]

Analogous lexical entries can be introduced for 'girl', 'boy', 'man', and 'male'.

The Default Position can then be understood as the conjunction of ('Woman': Gender Only) and ('Female': Sex Only). As we have seen, this combination of theses is not provisionally adequate: While it is consistent with INCLUSION, it is not consistent with DEFINITION because there is no necessary connection between being W-gendered and being F-sexed.[25]

Moreover, given that some trans women are of Sex M, no provisionally adequate theory can include ('Woman': Sex Only), for this would make incorrect predictions about INCLUSION.[26] Thus, the Sex-Only View — which combines ('Woman': Sex Only) with ('Female: Sex Only) — is also provisionally inadequate.

Indeed, if we restrict our attention to the four candidate lexical entries above, only one combination of views about 'woman' and 'female' is potentially provisionally adequate: the combination of ('Woman': Gender Only) with ('Female': Gender Only). This combination validates INCLUSION on the assumption, motivated above, that 'trans' is an intersective adjective, and it validates DEFINITION on the assumption that in order for an individual to be W-gendered, she must be human. Let us call this combination of views the *Gender-Only View.*

Although the Gender-Only View is provisionally adequate, however, it faces serious empirical difficulties. Whatever gender is, it is not the sort of thing which can be possessed by organisms like insects, fish, and plants. Yet it seems clear that we can truly describe such organisms using 'female'. Even when we restrict our attention to humans, it seems that there are individuals which it is appropriate to describe using 'female' but who are not W-gendered. For example, individuals who grow up outside of the context of any larger society may lack the social properties required to belong

---

[24] We can define these positions formally as follows:

> ('Woman': Gender Only): $[\![\text{woman}]\!] = \lambda x.x$ is a W-gendered adult.

> ('Woman': Sex Only): $[\![\text{woman}]\!] = \lambda x.x$ is an adult human of Sex F.

> ('Female': Gender Only): $[\![\text{female}]\!] = \lambda x.x$ is W-gendered.

> ('Female': Sex Only): $[\![\text{female}]\!] = \lambda x.x$ is of Sex F.

[25] More precisely, given EXISTENCE and DIVERSITY, not every W-gendered adult F-sexed, so the Default Position must hold that DEFINITION is false.

[26] Putting the same point a different way, no theory which endorses ('Woman': Sex Only) can account for our observation that 'trans' is not a non-subsective adjective.

to a gender category. Yet it seems clear that some such individuals can correctly be described using 'female'. For these reasons, although the Gender-Only View is provisionally adequate, it is not ultimately empirically defensible.

So, despite the inviting simplicity of candidate lexical entries according to which binary category terms pick out either sex or gender exclusively, it seems that there is no combination of such lexical entries which yields an overall semantic theory that is both provisionally adequate and empirically defensible. The semantics of 'woman' and 'female' is a more complex affair than might initially have been supposed.

4. Gender First

Problems arise for the Gender-Only View when we consider non-gendered individuals. A natural way to revise the Gender-Only View, then, is to preserve its predictions about the extensions of 'woman' and 'female' over the set of gendered individuals while revising its predictions about the extensions of 'woman' and 'female' over the set of non-gendered individuals so that they are sensitive to sex rather than gender. I have called views with this structure *gender-first* views, since they hold that binary category terms first "check" whether an individual is gendered and, if so, pick out that individual's gender.

Why think that a gender-first account of binary category terms might be on the right track? In addition to the fact that such an account can be made provisionally adequate while avoiding the problems with the Gender-Only View, there is evidence that related natural-language terms work in a structurally similar way.

Consider, for example, the meanings of genealogical terms like 'parent', 'mother', 'father', and so on. Binary category terms and genealogical terms are similar in that they exist at the intersection of social reality and underlying biology: in the case of binary category terms, the biological reality of sex and the social reality of gender; in the case of genealogical terms, the biological reality of reproduction and the social reality of parenting or legally recognized guardianship.

Just as one might be inclined toward a gender-only view or a sex-only view of binary category terms, one might be inclined toward a social-only view or a biological-only view of genealogical terms. According to a social-only view of 'parent' for example, that term might denote the property of being legally recognized as the guardian of one or more dependents. According to a biological-only view, on the other hand, 'parent' might denote the property of having one more offspring. Each of these theories would face difficulties parallel to the difficulties faced by ('Woman': Gender Only) and ('Woman': Sex Only): Against the social-only view, we might object that we very naturally apply genealogical terms to nonhuman animals; against the biological-only view, we might object that adoptive parents are parents in the fullest sense.

What seems to be going on in the genealogical case is that the meanings of our terms track social reality rather than the underlying biology wherever this social reality has been constructed (i.e. when describing human societies). When there are no social facts to track, however, they track biological relationships. In other words, genealogical terms seem to function in a *social-first* way. This means that they must pick out complex properties: to be a parent is to have a certain social property if one

is a human and a certain quite different biological property if one is not.[27] The proposal that binary category terms work in a gender-first way simply applies this more general social-first pattern to the case of gender and sex.

I suggest that the correct semantic theory of 'woman' and 'female' consists in the following two lexical entries:

> ('Woman': Gender First): 'woman' picks out the property of being either (i) a W-gendered adult human or (ii) a non-gendered F-sexed adult human.

> ('Female': Gender First): 'female' picks out the property of being either (i) a W-gendered human or (ii) an F-sexed individual who is not a gendered human. [28]

Let us call the combination of ('Woman': Gender First) and ('Female': Gender First) the *Gender-First View*.[29]

The Gender-First View validates both INCLUSION and DEFINITION. The validity of INCLUSION follows from the fact that trans women are W-gendered adults. The validity of DEFINITION can be established by reasoning from cases. Suppose that an individual falls in the extension of 'woman' in

---

[27] This is admittedly a simplification of the empirical facts. For example, we can say of a biological parent who has lost legal custody of their child that they are a mother or a father. There thus seems to be some ambiguity, polysemy, or context sensitivity associated with genealogical terms. Non-univocality of this kind is no objection to the point I am trying to make, however, since there is good evidence that at least one of the meanings of 'parent' must work in a social-first way. If this were not so, and 'parent' instead had only social-only and biological-only interpretations, we would (incorrectly) predict that it should be acceptable in some contexts to deny (e.g.) that a female sea turtle laying her eggs on the beach is a parent. This is because sea turtles do not have whatever properties are required to count as parents in the social sense. Positing ambiguity between a biological-only interpretation and a social-first interpretation of 'parent' avoids this prediction because it allows that nonhuman animals count as parents in virtue of their biological relationships in all contexts.

[28] We can define these positions formally as follows:

> ('Woman': Gender First): $[\![woman]\!] = \lambda x.x$ is a W-gendered adult human or (x is not gendered and x is an F-sexed adult human).

> ('Female': Gender First): $[\![female]\!] = \lambda x.x$ is a W-gendered human or (x is not a gendered human and x is F-sexed).

[29] The Gender-First View can be complicated slightly to avoid predicting that adult F-sexed non-binary individuals fall into the extension of 'woman' and 'female' in virtue of lacking a gender. A simple fix is to assign special importance to the property of lacking a gender in virtue of facts about one's identity (rather than, for example, because one has grown up outside human society). For example, we might revise ('Woman': Gender First) along these lines by replacing 'non-gendered' with 'lacking a gender for reasons other than gender identity'. It is worth noting, however, that some non-binary people describe themselves as belonging to a gender category rather than as not belonging to any gender category. See Dembroff (2020, 9) for discussion.

virtue of being a W-gendered adult human. Then that individual falls in the extension of 'female' in virtue of being a W-gendered human and the extension of 'adult human' in virtue of being an adult human. Suppose instead that an individual falls in the extension of 'woman' in virtue of being an F-sexed but not gendered adult human. Then that individual falls in the extension of 'female' in virtue of being F-sexed but not gendered and in the extension of 'adult human' in virtue of being an adult human. Similar reasoning can be used to establish the right-to-left direction of the equivalence between 'woman' and 'adult human female'.

At the same time, the Gender-First View escapes the empirical problems faced by the Gender-Only View. ('Female': Gender First) provides a straightforward explanation of why it is appropriate to classify certain nonhuman organisms as female. And the Gender-First View correctly predicts that individuals who exist outside the societies which construct gender categories can nevertheless be correctly described as women and females: such individuals lack genders and thus fall in the extension of both predicates in virtue of their sex.

Despite its empirical advantages, some might object to the Gender-First View on the basis that it holds that binary category terms pick out disjunctive properties. How, they might ask, could such familiar expressions as 'woman' and 'female', which we learn and deploy with little effort, pick out such complex properties? But humans are frequently able to learn and deploy complex linguistic rules. Consider again the analogy with 'parent'. If 'parent' works in a social-first way, then it picks out a property just as disjunctive as those which feature in ('Woman': Gender First) and ('Female': Gender First). But speakers have no trouble acquiring competence with 'parent'. So the complexity of the Gender-First View should not be held against it.[30]

5. Non-Univocality: Instability

Let's take stock. I have argued that an acceptable theory of binary category terms in inclusive idiolects should validate INCLUSION and DEFINITION, and also that the Gender-First View satisfies this condition while avoiding the empirical problems that face the Gender-Only View. Of the theories discussed so far, then, the Gender-First View seems to do best at capturing the linguistic data.

In this section and the next, I consider the possibility that binary category terms are context-sensitive, ambiguous, and/or polysemous (I will use *non-univocal* as a label which is neutral between these options). This is a possibility that has been raised in the literature, both with respect to 'woman' (Saul (2012), Barnes (2020)) and with respect to 'female' (Heartsilver (2021)). There are, broadly, two motivations for positing non-univocality in the meanings of binary category terms. First, there is a general skepticism that these words have stable meanings across English speakers. Second, there is data about particular uses of binary category terms which suggests that the same word can make a different truth-conditional contribution on different occasions. I discuss the first motivation in this section and the second in the next.

A recent example of the first kind of argument for non-univocality is Barnes's (2020) discussion of 'woman'. Barnes writes:

---

[30] Kirk-Giannini and Glanzberg (fc) make a similar point about the complex rules which speakers of languages that mark nouns for grammatical gender must internalize.

"Common usage of gender terms is currently in flux… Teenagers probably mean something quite a bit different from their grandparents; wealthy teenagers in Manhattan probably mean something quite a bit different from working class teenagers in Alabama." (2020, 712)

On this basis, she concludes that:

"…our gender terms are complex, messy, and often refer to a gerrymandered cluster of features — including sex, perceived sex, gender identity, gender expression, etc. They can and do mean different things in different contexts. I don't think there's any one uniquely correct definition of terms like 'woman', and definitional projects which seek to give necessary and sufficient conditions for the application of terms like 'woman' seem doomed to counterexamples." (2020, 719-20)

This leads her to endorse a radically context-sensitive semantics for binary category terms, according to which they can express any of a very wide range of properties depending on the context.[31]

A related worry about the stability of the meanings of binary category terms is that everyday English speakers may not draw a firm conceptual distinction between sex and gender. Mikkola (2022), for example, notes that "most ordinary language users appear to treat [sex and gender] interchangeably."

These sorts of worries do not provide strong evidence for the non-univocality of binary category terms, however. While it may indeed be that various communities use binary category terms differently, this observation only suggests that we may need separate semantic theories to describe the meanings of binary category terms in different communities. Compare: The fact that 'pants' means only *underpants* in British English and only *trousers* in American English does not indicate that 'pants' is ambiguous or context sensitive in either language community. Once we have focused in on a language community and begun to investigate the meanings it assigns to binary category terms, there is little reason to doubt the validity of the empirical methods of natural language semantics, on which my arguments for the Gender-First View are based.

Similarly, one does not need to conceptually distinguish between sex and gender to use binary category terms in a gender-first way. All that is required is that one's applications of the terms be counterfactually sensitive to gender in the relevant cases and to sex in others. Compare: One need not conceptually distinguish between common law marriage and traditional marriage in order to use the word 'married' to pick out the disjunctive property of having either participated in a legally recognized marriage ceremony or met the requirements for a common law marriage. All that is required is that one's applications of the word 'married' pick out individuals who have this disjunctive property and exclude others.

So there is little reason to accept the non-univocality of binary category terms on the basis of a general skepticism about whether they have stable meanings.

6. Non-Univocality: Context Sensitivity

---

[31] A similar proposal is described in Saul (2012). Saul ultimately rejects it, however, on the basis that it does not secure the truth of 'Trans women are women' in all or even most contexts.

Setting aside general skepticism about the stability of the meanings of binary category terms, there is considerable evidence that at least 'female' and 'male' have non-univocal meanings in everyday English. We have already considered the evidence that 'female' has a meaning which works as described by the Gender-First View. But there do seem to be contexts in which 'female' functions as a sex term. For example, it certainly appears that 'female' in phrases like 'female reproductive anatomy' is correctly described by ('Female': Sex Only). Indeed, it is likely attention to examples of this sort which has attracted so many theorists, regardless of their positions on the semantics of 'woman', to the view that 'female' is a sex term.

One might attempt to resist the idea that 'female' ever functions as a sex term. Perhaps phrases like 'female reproductive anatomy' pick out the reproductive anatomy of F-sexed individuals because they pick out whatever reproductive anatomy is typical or statistically normal for W-gendered individuals, and most W-gendered individuals are F-sexed. But it is difficult to see how this strategy could work in full generality. Imagine, for example, that a doctor instructs her office staff using (9):

> (9) Please arrange a cervical cancer screening appointment for each female patient over the age of fifty.

Even if we imagine that both her W-gendered and her M-gendered patients are diverse with respect to sex, (9) is felicitous, and its natural interpretation involves booking appointments for all and only the F-sexed patients over the age of fifty. Better, then, to treat 'female' as non-univocal, expressing in most contexts a gender-first meaning but capable in some contexts of expressing a sex-only meaning. An analogous point can be made about 'male'.

Do similar cases suggest non-univocality for 'woman'? Saul (2012, 200) offers some examples which she takes to suggest this conclusion. She writes:

> "I have no problems with forensic scientists uttering sentence ([10]).
>
> > ([10]) This is the DNA of a woman.
>
> And this is not because I think DNA is a very good, though fallible, guide to who is likely to live out a woman's gender role. I would also assent to ([11]), with "woman" used as a sex term.
>
> > ([11]) It is important for scientists testing drugs to study both men and women."

I do not think examples of this kind provide strong support for the conclusion that 'woman' is non-univocal. It is methodologically important to distinguish between judgments about the truth or assertability of sentences, which constitute the primary source of evidence for semantic theorizing, and judgments about *why* certain sentences are true or assertable. So, while I share Saul's judgments that (10) and (11) are often assertable, I do not think there is good evidence for her claims that (10)'s assertability does not have to do with gender and that 'woman' in (11) is naturally used as a sex term.

In order to assess whether (10) and (11) are good evidence for non-univocality in 'woman', we should imagine a scenario in which it is salient to us that not all W-gendered individuals are F-sexed and not all F-sexed individuals are W-gendered — cases where every W-gendered individual is F-

17

sexed and every F-sexed individual is W-gendered are ones in which 'woman' will always pick out the same individuals whether or not it sometimes functions as a sex term. Suppose, for example, that the forensic scientist who utters (10) is providing information about who committed a crime, and that the suspects prior to her utterance of (10) include W-gendered individuals of both sexes and M-gendered individuals of both sexes.

According to Saul, in such circumstances (10) ought to have a reading on which it is perfectly felicitous and exonerates all suspects who are either cis men or trans women. My own intuition, however, is that — while we could certainly understand what the speaker of (10) is trying to get at because she makes it clear that what epistemically grounds her assertion is a fact about DNA — her utterance is markedly less felicitous than if she had instead said "This is the DNA of a female." This is plausibly because in situations where sex and gender are not correlated, DNA is not a reliable indicator of gender. So (10) does not offer strong support for the claim that 'woman' is non-univocal.

A similar point can be made about (11), which according to Saul can express the proposition that it is important for scientists testing drugs to study adult humans of both Sex M and Sex F. If this interpretation is available, then it should be felicitous to describe a situation in which a drug is tested exclusively on cis women and trans women as one in which it is "tested on both men and women." But such a way of characterizing the situation in question would at best be seriously misleading. In general, it seems that sentences like (10) and (11) are most felicitous when they are used in situations where it can be assumed that an individual is W-gendered just in case she is F-sexed, and so do not constitute a strong case for the non-univocality of 'woman'.[32]

Returning to 'female', I will operate on the assumption that its non-univocality is a form of context sensitivity. It is preferable to posit lexical ambiguity only in the presence of evidence that the different lexical entries associated with a word are assigned to distinct lexical items in different languages, and there does not appear to be any such evidence when it comes to 'female' or 'male'.[33] As between treating the non-univocality as polysemy or context sensitivity, I will plump for context sensitivity, though I see no reason why an analysis in terms of polysemy could not succeed. Amending ('Female': Gender First) to incorporate this context sensitivity gives us:

> ('Female': Gender First [Context Sensitive]): Depending on context, 'female' picks out either (i) the property of (being a W-gendered human or (being F-sexed and not a gendered human)) or (ii) the property of being F-sexed.

---

[32] Some may not share my intuitions about (10) and (11), insisting that 'woman', like 'female', is context sensitive. To them I offer the reply I gave in Section 2 to those who reject INCLUSION or DEFINITION: my project is to provide a semantic theory of binary category terms as they are used in a specific class of idiolects, which I take to be spoken by a significant proportion of the population. In these idiolects, I take it that 'woman' is not context sensitive in the way Saul suggests. But my project should be of interest even to those who do not share the linguistic intuitions that guide it, since it shows how it is possible to assign meanings to 'woman' and 'female' on which INCONSISTENCY is false.

[33] See Sennet (2021) for a description of other tests for lexical ambiguity, which also suggest that 'female' is not ambiguous.

We can then understand the Gender-First View as the conjunction of ('Woman': Gender First) and ('Female': Gender First [Context Sensitive]).

Once we have admitted that 'female' can in certain contexts function as a sex term, we might be tempted to account for its linguistic behavior by treating it as context sensitive between a gender-only meaning and a sex-only meaning rather than between a gender-first meaning and a sex-only meaning, as I have suggested. But this temptation should be resisted. Because the resulting theory would hold that 'female' expresses a gender-only meaning in some contexts, it would incorrectly predict that it is acceptable in some contexts to deny that (e.g.) a pregnant cat is female, since no cat is W-gendered. The two meanings of 'female' can diverge only over gendered individuals.[34]

How does the observation that 'female' is context sensitive affect the force of my arguments above? Until now, I have moved freely between descriptive and metalinguistic glosses of DEFINITION: I have not needed to distinguish between the claim that women are adult human females and the claim that 'Women are adult human females' expresses a truth. If 'female' is context sensitive, however, it becomes necessary to keep careful track of the descriptive–metalinguistic distinction. Let us therefore understand DEFINITION as the claim that the sentence 'Women are adult human females' expresses a truth. The context sensitivity of 'female' means that the truth of DEFINITION must be relativized to a context. INCLUSION and INCONSISTENCY should similarly be understood as metalinguistic claims whose truth is relativized to a context: INCLUSION is the claim that 'Trans women are women' expresses a truth, and INCONSISTENCY is the claim that 'Trans women are women' and 'Women are adult human females' express inconsistent contents.[35]

If 'woman' is not context sensitive, the evidence favoring INCLUSION suggests that it is true relative to all contexts.[36] Given that 'female' can function as a sex term in some contexts, however, DEFINITION is not true relative to all contexts. And in contexts where DEFINITION is false, INCONSISTENCY is true, since the truth of 'Women are adult human females' requires that 'woman' functions as a sex term while the truth of 'Trans women are women' requires that it does not.

The fact that INCONSISTENCY is true relative to some contexts does not undermine the dialectical significance of the argument of Sections 1 to 4. The central motivation for developing the Gender-First View was the observation that there is linguistic evidence for both INCLUSION and DEFINITION, and that a theory according to which both principles are true has important practical and theoretical benefits. Given that INCLUSION is true relative to all contexts, the linguistic evidence supporting DEFINITION is evidence that in many contexts the two principles can be true together, and thus that there are contexts relative to which INCONSISTENCY is false. Let us refer to such contexts as *consistent contexts*. I believe the linguistic evidence suggests that most contexts are consistent contexts; contexts in which 'female' functions as a sex term are special and must be induced by cues like a presupposed relationship between falling in the extension of 'female' and being at risk of cervical cancer, as in (9). And if most contexts are consistent contexts, this fact both

---

[34] This argument is structurally similar to the one offered in footnote 27 regarding 'parent'.

[35] Similarly, EXISTENCE is the claim that 'There are trans women' expresses a truth, and DIVERSITY is the claim that some individuals falling in the extension of 'trans woman' are not F-sexed.

[36] Since the meaning of 'woman' is not context sensitive, EXISTENCE and DIVERSITY are likewise true relative to all contexts.

explains the language trans individuals use to refer to themselves and helps to defuse 'Women are adult human females' as a trans-exclusionary slogan.

7. Objections and Replies

I have argued that there is much to recommend the Gender-First View as compared to the Default Position, the Gender-Only View, and the Sex-Only View. But the possibility remains that there might be some direct argument against the Gender-First View. In this section, I consider this possibility.

Because the Gender-First View consists in the conjunction of ('Woman': Gender First) and ('Female': Gender First [Context Sensitive]), any argument against it must be an argument targeting one or both of these principles. I have already addressed arguments for rejecting both principles in favor of a radically context-sensitive approach to binary category terms and found them unconvincing. So I take it that the most pressing worry for the Gender-First View is that there is some compelling argument for a gender-only or a sex-only semantics for 'woman' or 'female'.

We have seen that there is little to recommend ('Female': Gender Only), and I will proceed on the assumption that it cannot serve as the basis for a compelling objection to the Gender-First View. At the same time, arguments which might be thought to support ('Woman': Gender Only) generally turn on intuitions about which gendered adult humans count as women. Since ('Woman': Gender First) makes the same predictions about these cases as ('Woman': Gender Only), such arguments do not actually support ('Woman': Gender Only) as opposed to ('Woman': Gender First). A related point can be made about ('Female': Sex Only) — arguments which show that there are some contexts in which 'female' functions as a sex term are consistent with ('Female': Gender First [Context Sensitive]) and therefore do not constitute evidence against it.

To my mind, then, the most serious threat to the Gender-First View comes from the possibility that there is some direct argument for ('Woman': Sex Only). And indeed, a number of arguments for ('Woman': Sex Only) have recently been suggested in the literature. In the remainder of this section, I will focus on three arguments discussed in an exchange between Alex Byrne (2020, 2022) and Maggie Heartsilver (2021), expanding on Heartsilver's remarks to explain why I do not believe those arguments threaten the Gender-First View.

Byrne (2020) defends a thesis which he states as follows:

(AHF) Necessarily, for all S, S is a woman iff S is an adult human female.

(AHF) superficially resembles the thesis I have called DEFINITION. In fact, however, the two theses should be carefully distinguished. DEFINITION is a claim about the truth of a certain English sentence and is therefore neutral concerning the meanings of 'woman' and 'female'. In contrast, because Byrne assumes the truth of ('Female': Sex Only), in his idiolect (AHF) does not express a claim which is neutral concerning the meaning of 'woman' or 'female'. Instead, it expresses the proposition that S is a woman just in case S is an adult human of Sex F — that is, it expresses a disquoted version of ('Woman': Sex Only).

Byrne offers an array of arguments in support of (AHF). It is beyond the scope of my discussion here to respond to them all in detail. It is worth mentioning, however, that many are arguments for

DEFINITION rather than (AHF) — indeed, Byrne makes basically the same points that were made in support of DEFINITION in Section 2.[37] I will not discuss these arguments further, since they do not threaten the Gender-First View.

Byrne makes two epistemic arguments for (AHF) which, if they are compelling, support the truth of ('Woman': Sex Only) and therefore suggest the falsity of the Gender-First View. Both involve knowledge that an individual is a woman or girl in the apparent absence of any evidence other than that they are of Sex F.

The first argument concerns Mitochondrial Eve, an F-sexed human who lived around 100,000 years ago and passed on her mitochondrial DNA to all living humans. Byrne claims, and offers some linguistic evidence, that it is acceptable to refer to Mitochondrial Eve as a woman. (AHF) offers a simple explanation of this fact: we are in a position to know that Mitochondrial Eve was a woman because we are in a position to know that she was an F-sexed adult human, and that suffices for her to have been a woman.

The second argument concerns human infants. As Byrne points out, (AHF) is plausible if and only if the analogous principle about girls is plausible, namely:

    (JHF) Necessarily, for all S, S is a girl iff S is a juvenile human female.

Moreover, Byrne argues, (JHF) is supported by the fact that it is possible to know that an infant is a girl "by inspection" — that is, knowledge that an infant is F-sexed appears to suffice for knowledge that it is a girl.

Heartsilver (2021) suggests a similar response to both arguments: we are in a position to know that Mitochondrial Eve was a woman, and to know of an infant that it is a girl, only if we are in a position to know that they are not trans. If the friend of (AHF) were to insist that we can know that Mitochondrial Eve was a woman or that an infant is a girl *solely* on the basis of information about sex, she continues, this would be begging the question — the move would have to be supported by appeal to something like the principle that trans men are women, and this is not a principle that we have reason independent of (AHF) to accept. I find Heartsilver's reasoning here convincing, and so I think Byrne's arguments do not ultimately threaten the Gender-First View. But it is worth spelling out in more detail what a proponent of the Gender-First View might say about Byrne's cases.

The Gender-First View is a view about the meanings of binary category terms, not a view about the metaphysics of gender. For this reason, while I have assumed a distinction between W-gendered and M-gendered individuals, I have said little about what it takes for an individual to be W- or M-gendered. While it is beyond the scope of my discussion here to defend any particular account of gender, contemporary work on the metaphysics of gender suggests that a complete theory will hold that an individual's gender is determined by some combination of their social position (that is, how others' beliefs about their sex shape the way in which they are integrated into the social world) and their gender identity (that is, their own internal sense of themself as a man, a woman, or neither).

---

[37] Heartsilver (2021) emphasizes this feature of Byrne's arguments.

With respect to Mitochondrial Eve, either the community in which Eve lived was complex enough to have gendered social roles, or it was not. If it was not, then the Gender-First View predicts that Eve was a woman in virtue of being an F-sexed adult human. If it was, then it is conceptually possible that Eve occupied the social position of a man or had a male gender identity. This possibility seems rather unlikely in the context of an isolated society of hunter-gatherers — but it is a possibility.[38] And so, combined with plausible premises about knowledge and the norms on assertion, the Gender-First View predicts that sentences like 'Mitochondrial Eve was a woman' should strike speakers as acceptable when the possibility that Mitochondrial Eve was not W-gendered is not salient and less acceptable when it is. This prediction appears to be correct: in normal contexts, it is unproblematic to refer to Mitochondrial Eve as a woman, but when the possibility that Mitochondrial Eve was not W-gendered is raised to salience, the assertability of sentences like 'Mitochondrial Eve was a woman' is degraded.

To assess the argument concerning infants, we need a gender-first lexical entry for 'girl'. The natural option is the following:

> ('Girl': Gender First): 'girl' picks out the property of being either (i) a W-gendered juvenile human or (ii) a non-gendered F-sexed juvenile human.[39]

It seems clear that infants are not cognitively developed enough to have a gender identity. So either they are gendered in virtue of their social positions or they are not gendered. If they are not gendered, an F-sexed infant counts as a girl in virtue of being an F-sexed juvenile human. If they are gendered, an F-sexed infant counts as a girl in virtue of being assigned a certain social role based on others' beliefs that she is F-sexed. Either way, the Gender-First View is consistent with the idea that infants can be correctly classified as girls or boys "by inspection."

In her response to Byrne, Heartsilver suggests that 'female' is sometimes used in a way which includes trans women and some intersex individuals. The Gender-First View is one way of developing this idea. Byrne (2022), however, believes Heartsilver's suggestion should be rejected for empirical reasons. The third argument I will consider in this section, then, is Byrne's response to Heartsilver.

Byrne interprets Heartsilver as holding that 'female' sometimes picks out the property of *identifying as female*. He argues that this suggestion makes the wrong prediction about certain cases. For example, he asks us to "consider a (nontransgender) butch lesbian who is unbothered by frequent 'misgendering,' routinely lies when forms ask for her gender, does not feel part of the female club, and who embraces her masculinity," and concludes that "if 'woman' did have a sense corresponding to 'female'-as-used-to-mean-*identifies as female*, it would not apply to this person" (2022, 11) This result is a problem because "there is no everyday sense of 'woman' in which some lesbians are not women" (2022, 11).

---

[38] Heartsilver (2021, 9) makes a similar point.

[39] More formally, we have:

> ('Girl': Gender First): $[\![girl]\!] = \lambda x.x$ is a W-gendered juvenile human or (x is not gendered and x is an F-sexed juvenile human).

This objection need not trouble the proponent of the Gender-First View. The Gender-First View does not take a position on what it takes for an individual to be W-gendered and is therefore not committed to the idea that gender category membership is determined exclusively by gender identity in the way Byrne suggests. In any case, it is a desideratum on any adequate theory of gender identity that it does not predict that in Byrne's case the individual in question does not identify as a woman — in her diary, for example, she surely refers to herself as a woman. At best, what Byrne's case shows is not that Heartsilver's suggestion about 'female' (and thus also the Gender-First View) is untenable, but rather that butch lesbians may constitute an important test case for theories of gender identity.

8. Conclusion

The Gender-First View opens up the possibility of rejecting INCONSISTENCY by holding that both 'Trans women are women' and 'Women are adult human females' can express truths in the same context. In so doing, it captures the empirical data supporting INCLUSION and DEFINITION and makes sense of the existence and prevalence of consistent idiolects. As a theory of such idiolects, it is empirically superior to the Sex-Only View because it validates INCLUSION in all contexts and to the Default Position because it validates DEFINITION in most contexts. It also has advantages over radically context-sensitive approaches in so far as it is simpler, validates INCLUSION in all contexts and DEFINITION in most contexts, and explains how INCONSISTENCY could fail in any context. At the same time, the Gender-First View allows us to understand the way trans individuals talk about themselves as continuous with the linguistic practice of ordinary speakers, and it helps to disarm the sentence 'Women are adult human females' as a rhetorical tool for anti-trans activism.

Works Cited

Ásta. (2011). The Metaphysics of Sex and Gender. In C. Witt (ed.) *Feminist Metaphysics*, Springer, pp. 47–65.

Ásta. (2013). The Social Construction of Human Kinds. *Hypatia* 28: 716–732.

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

Bach, T. (2012). Gender is a Natural Kind with a Historical Essence. *Ethics* 122: 231–272.

Barnes, E. (2020). Gender and Gender Terms. *Noûs* 54: 704–730.

Barnes, E. (2022). Gender without Gender Identity: The Case of Cognitive Disability. *Mind* 131: 836–862.

Bettcher, T. M. (2009). Trans identities and first-person authority. In L. Shrage (Ed.), *"You've Changed": Sex Reassignment and Personal Identity*. Oxford University Press, pp. 98–120.

Butler, J. (1990). *Gender Trouble*. Routledge.

Butler, J. (1993). *Bodies that Matter*. Routledge.

Byrne, A. (2020). Are Women Adult Human Females? *Philosophical Studies* 177: 3783–3803.

Byrne, A. (2021). Gender Muddle: Reply to Dembroff. *Journal of Controversial Ideas* 1(5): 1–24.

Byrne, A. (2022). The Female of the Species: Reply to Heartsilver. *Journal of Controversial Ideas* 2(11): 1–22.

Del Pinal, G. (2015). Dual Content Semantics, Privative Adjectives, and Dynamic Compositionality. *Semantics & Pragmatics* 8(7): 1–53.

Dembroff, R. (2016). What is Sexual Orientation? *Philosophers' Imprint* 16(3): 1–27.

Dembroff, R. (2020). Beyond Binary: Genderqueer and a Critical Gender Kind. *Philosophers' Imprint* 20(9): 1–23.

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

Fausto-Sterling, A. (1993). The Five Sexes: Why Male and Female Are Not Enough. *The Sciences* 33: 20–24.

Fausto-Sterling, A. (2000). *Sexing the Body*. Basic Books.

George, B. R. and Briggs, R. A. (2019). Science Fiction Double Feature: Trans Liberation on Twin Earth. (Unpublished Manuscript).

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

Heartsilver, M. (2021). Deflating Byrne's "Are Women Adult Human Females?" *Journal of Controversial Ideas* 1(9): 1-16.

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

Kamp, H. (1975). Two Theories About Adjectives. In E. Keenan (ed.), *Formal Semantics of Natural Languages*, Cambridge University Press, pp. 123–155.

Kamp, H. and Partee, B. (1995). Prototype Theory and Compositionality. *Cognition* 57: 129–191.

Kaur, H. (2022, December 15). Cambridge Dictionary's Definitions for 'Man' and 'Woman' Now Include Trans People. *CNN*. <https://www.cnn.com/2022/12/15/us/cambridge-dictionary-woman-definition-trans-cec/index.html>

Kirk-Giannini, C. D. and Glanzberg, M. (fc). Pronouns and Gender. In Anderson, L. and Lepore, E. (eds.) *The Oxford Handbook of Applied Philosophy of Language*.

Lawford-Smith, H. (2022). *Gender-Critical Feminism*. Oxford University Press.

McKitrick, J. (2015). A Dispositional Account of Gender. *Philosophical Studies* 172: 2575–2589.

Mikkola, M. (2016). *The Wrong of Injustice: Dehumanization and Its Role in Feminist Philosophy.* Oxford University Press.

Mikkola, M. (2022). Feminist Perspectives on Sex and Gender. In Zalta, E.N. and Nodelman, U. (eds.) *The Stanford Encyclopedia of Philosophy* (Winter 2022 Edition). <https://plato.stanford.edu/entries/feminism-gender/>.

Partee, B. (2007). Compositionality and Coercion in Semantics: The Dynamics of Adjective Meaning. In Bouma, G., Krämer, I., and Zwarts, J. (eds.) *Cognitive Foundations of Interpretation*, Royal Netherlands Academy of Arts and Sciences Press, pp. 145–161.

Rubin, G. (1975). The Traffic in Women: Notes on the "Political Economy" of Sex. In R. Reiter (ed.) *Toward an Anthropology of Women*, Monthly Review Press, pp. 157–210.

Saul, J. (2012). Politically Significant Terms and the Philosophy of Language: Methodological Issues. In S. L. Crasnow and A. M. Superson (eds.) *Out from the Shadows: Analytical Feminist Contributions to Traditional Philosophy*, Oxford University Press, pp. 195–216.

Sennet, A. (2021). Ambiguity. In Zalta, E.N. (ed.) *The Stanford Encyclopedia of Philosophy* (Fall 2021 Edition). <https://plato.stanford.edu/archives/fall2021/entries/ambiguity/>.

Stoljar, N. (1995). Essence, Identity, and the Concept of Woman. *Philosophical Topics* 23: 261–293.

Stoljar, N. (2011). Different Women: Gender and the Realism–Nominalism Debate. In C. Witt (ed.) *Feminist Metaphysics*, Springer, pp. 27–46.

Witt, C. (2011a). *The Metaphysics of Gender.* Oxford University Press.

Witt, C. (2011b). What is Gender Essentialism? In C. Witt (ed.) *Feminist Metaphysics*, Springer, pp. 11–25.

Wynn, N. (2019). Are Traps Gay? YouTube video transcript. <https://www.contrapoints.com/transcripts/are-traps-gay?>

Zimman, L. (2014). The Discursive Construction of Sex: Remaking and Reclaiming the Gendered Body in Talk About Genitals Among Trans Men. In Zimman, L., Davis, J. L., and Raclaw, J. (eds.) *Queer Excursions: Retheorizing Binaries in Language, Gender, and Sexuality*, Oxford University Press, pp. 13–34.