Spring 5-4-2022

Self-identification

Maximiliana Jewett Rifkin

Recommended Citation
Rifkin, Maximiliana Jewett, "Self-identification." Thesis, Georgia State University, 2022.
doi: https://doi.org/10.57709/28907105

Self-Identification

by

Maximiliana Jewett Rifkin

Under the Direction of Neil Van Leeuwen, PhD

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2022

ABSTRACT

Here, I first analyze gender identity qua gender self-ascription and offer a theory of the psychological states underpinning gender self-ascriptions, which I call a form of 'self-identification'. I hold gender self-identification consists of a gender self-concept, which itself consists of a belief or assumption in a context, and sometimes involves a gender role ideal, which consists of an individual's expectations and standards for how to perform a gender role. Second, I defend my view from an objection to similar views like it, which amounts to the claim that they cannot explain nonbinary gender identities, by showing how my account explains various nonbinary gender identities in psychological terms. Finally, I show how my view can begin stabilizing the construct of gender identity in neuroscience by helping researchers foster agreement on how to define gender identity and what methods are best to study it.

INDEX WORDS: Gender identity, self-identification, construct stability, construct validity, interdisciplinary

Self-Identification

by

Maximiliana Jewett Rifkin

Committee Chair: Neil Van Leeuwen

Committee: Christie Hartley

Dan Weiskopf

Electronic Version Approved:

Office of Graduate Services

College of Arts and Sciences

Georgia State University

May 2022

DEDICATION

I dedicate this thesis to my family and friends who have stalwartly weathered the trials of my trans life beside me by offering the support, encouragement, and insights I needed to carry on. The road to this text was paved in part by their generosity and kindness.

ACKNOWLEDGEMENTS

I would like to acknowledge Neil Van Leeuwen, Christie Hartley, and Dan Weiskopf with the utmost gratitude for their superb and indefatigable guidance as instructors, mentors, and colleagues.

TABLE OF CONTENTS

1 INTRODUCTION

Consider trans[1] women. We self-ascribe a feminine gender, and we experience ourselves as women. So it is with cis women, who also self-ascribe a feminine gender and experience themselves as women; and with trans and cis men, who self-ascribe a masculine gender and experience themselves as men. But now consider agender people. They do not self-ascribe any gender, and instead experience themselves as people without one. What can explain the former group's various gender self-ascriptions on the one hand, and agender people's lack of gender self-ascriptions on the other? To answer this question, we need an account of the psychological states that people use to self-ascribe a gender. Here, I offer an account of these psychological states, and I call what they realize a form of 'self-identification'[2] I think is specific to gender self-ascriptions.

Before we discuss my project in detail, however, we should contrast it with investigations of gender performed by theorists with different disciplinary backgrounds and interests. Hitherto, linguists have studied the meaning of gender terms like 'man' and 'woman' (McElhinny, 2003; Zeman, 2020), sociologists have investigated the societal imposition of these selfsame gender categories (Orit & Irby, 2017), and metaphysicians have analyzed how and whether we should construct gendered social categories in the first place (e.g., Bettcher, 2009; Jenkins, 2016, 2018; Dembroff, 2020). For ease of reference, let's call what linguists are interested in the 'semantics of gender'; what sociologists fancy knowing about, the 'social roles of gender'; and what metaphysicians focus on 'the ontology of gender'. My project, in contrast, is bound up with what

[1] Throughout this article, I will be using 'trans' as a standalone term to indicate the class of identities corresponding to 'trans man' and 'trans woman' rather than the umbrella term indexing other identities such as 'drag king and queen'.

[2] Although 'self-identification' is used to denote a variety of self-ascribed identities, there seems to me to be no obvious reason to assume that every self-identification works the same way, so I intend my use of the term to cover only *gender self-ascriptions*. Nevertheless, my restricted theory of self-identification may inform theories of the phenomenon that cover other instances.

I call the ‘psychology of gender’ and is best associated with the philosophy of mind. I aim to analyze the nature of the self-identification manifest when people self-ascribe a gender, whether feminine, masculine, or otherwise, and this is what I will mean by ‘gender identity’ throughout this article.

Contrast established, let’s turn to the details of my proposal. My project concerns Katherine Jenkins’s ‘norm relevancy account’, which Jenkins originally proposed as an account of prescriptive gender ontology establishing how we should construct our gender categories of ‘man’, ‘woman’, and ‘nonbinary’. Here, I am going to adapt the norm relevancy account to explain gender identity in the ‘psychology of gender’ sense introduced above, which targets only the self-identification manifest by gender self-ascriptions. I invoke different criteria for self-identification with which I aim to capture the different psychological routes to gender self-identification. On my account, then, someone self-identifies as a gender ‘x’ if they satisfy certain psychological criteria like having the belief or context-based assumption that they are ‘x’. More criteria can be added to supplement this analysis, and thus the current project is best thought of as a starting point for an empirical research program.

Despite my clarifications about my theoretical goals, my adaptation of the norm relevancy account to gender psychology may prompt worries that Robin Dembroff’s (2020) objection to the norm relevancy account will apply to my revised and repurposed version. However, my adaptation promises to resolve the objection Dembroff raises for the norm relevancy account when it is applied to gender ontology and psychology alike. Contrary to what Dembroff implies, the norm relevancy account helps rather than hinders our understanding of nonbinary gender identities.

In what follows, section two will summarize Jenkins’s norm relevancy account. Afterward, section three will show how the norm relevancy account can be adapted to explain gender identity qua gender psychology. Section four will then summarize Dembroff's objection to the norm relevancy account, and section five will explain why my version of the account evades their objection when applied to psychological and ameliorative gender alike. Finally, section six will discuss how my account of gender identity promises to help stabilize the construct of gender identity in neuroscience.

2 JENKINS, THE NORM RELEVANCY ACCOUNT, & AMELIORATIVE GENDER

Katherine Jenkins originally proposed the norm relevancy account to specify gender identity in the *ameliorative gender* sense, which as we saw above concerns how we should construct gendered social categories. In her work on the norm relevancy account, Jenkins (2016, 2018) has argued it addresses two central problems regarding gender identity in this sense: 1) an inclusion problem in a previous ameliorative analysis of the gender term 'woman' by Sally Haslanger, and 2) a circularity problem in trans activists use of the term 'gender identity' to define gender terms. Jenkins argues the inclusion problem arises from Haslanger's (2012) social position account of ameliorative gender – what Jenkins calls 'gender as class' – because it excludes some trans women from counting as women. In contrast, Jenkins argues the circularity problem arises because trans activists often define both gender terms and gender identity as 'a sense of oneself as a man, woman, or some other gender' and then use the latter to define the former. Jenkins attempts to solve these problems with the norm relevancy account, which builds on Haslanger's social position account by developing what Jenkins calls 'gender as identity'. The norm relevancy account thus specifies ameliorative gender using Jenkins's conception of 'gender as identity', which she takes to be partly determined in an important way by gender as class. To understand the norm relevancy account, then, we first need to review Jenkins's definitions of 'gender as class' and 'gender as identity'.

Jenkins begins her two-pronged account of ameliorative gender with the observation that insofar as "genders are subject positions that emerge from . . . a complex social matrix of practices, norms, institutions, material structures, rationales, and so forth . . . governed by a dominant [gender] ideology, [these subject positions] can be lived from within in ways that depart from, and may even run counter to, the logic of the system within which [they] developed" (Jenkins, 2016, pp. 407-408). Accordingly, she argues there are "two aspects of the matrix of practices that constitutes gender to which we need to be able to refer" and uses "the term 'gender as class' to refer to the way that gendered subject positions are defined by dominant [gender] ideology and the term 'gender as identity' to refer to the way that gendered subject positions are taken up by individuals" (Jenkins, 2016, p. 408).

Next, Jenkins bases her definition of 'gender as class' on Haslanger's social position account of gender, which analyzes gender in terms of the external, social feature of being perceived by others as a woman or man. For Haslanger, this perception is mediated by someone 'functioning as a woman' or 'functioning as a man' in a context. On Haslanger's account, then, someone is a woman if and only if they regularly and for the most part function as a woman in some context. For Haslanger, someone 'functions as a woman in a context C' if and only if[3]:

(i) S is observed or imagined in C to have certain bodily features presumed to be evidence of a female's biological role in reproduction;

(ii) That S has those features marks S within the background ideology of C as someone who ought to occupy certain kinds of social positions that are in fact subordinate (and so motivates and justifies S's occupying such a position); and

[3] Haslanger also offers a counterpart definition of 'functioning as a man in a context C' which defines men as individuals "who are *privileged* on the basis of observed or imagined bodily features presumed to be evidence of a *male* role in biological reproduction" (Jenkins, 2016, p. 398).

(iii) The fact that S satisfies (i) and (ii) plays a role in S's systematic subordination in C, that is, *along some dimension*, S's social position in C is oppressive, and S's satisfying (i) and (ii) plays a role in that dimension of subordination.

Accordingly, Jenkins defines 'gender as class' using "condensed versions of Haslanger's proposed target concepts of gender" such that:

(i) S is classed as a woman within context C iff S is marked in C as a target for subordination on the basis of actual or imagined bodily features presumed to be evidence of a female's role in biological reproduction.

Whereas:

(ii) S is classed as a man within context C iff S is marked in C as a recipient of privilege based on the actual or imagined bodily features presumed to be evidence of a male's role in biological reproduction.

Finally, Jenkins defines 'gender as identity' in relation to 'gender as class'. Whereas on Jenkins's account 'gender as class' constitutes the objective social matrices of gender norms (among other things) which instantiate the gendered subject positions individuals occupy, 'gender as identity' constitutes the subjective sense of personal relevance individuals feel for gender norms in virtue of occupying them. First, Jenkins bases this idea of having a subjective sense of personal relevance for gender norms on Haslanger's conception of racial identity as "an embodied phenomenon" with "important components . . . that are somatic, largely habitual, regularly unconscious, and often ritualized" (Jenkins, 2016, p. 9). Second, Jenkins recruits Haslanger's analogy that "one's identity is a . . . map that functions in a multitude of ways to guide and direct exchanges with one's social and material realities . . . which may be sometimes tacit and unconscious and sometimes explicit and conscious" (Jenkins, 2016, p. 9). Third and last, Jenkins thinks these ‘embodied, internal maps’ explain the variance between the gender norms prescribed for someone in a gendered subject position on the one hand and someone’s gendered behavior stemming from their subjective sense of relevance for those norms on the other, such that “a map could [either] guide someone classed as a woman . . . toward behaviors that are prescribed as ‘feminine’ by dominant ideology” or “to resist norms of acceptably feminine behavior” (Jenkins, 2016, pp. 412-413).

For Jenkins, then, ‘gender as identity’ is an embodied, internal map that may be either tacit, as when experiencing bodily feelings like anxiety while speaking as a woman in a male dominated space, or explicit, as when situations like this spur thoughts about whether to conform to the norm that women shouldn’t speak in some spaces. By prompting affects and thoughts like these, this embodied, internal map guides and directs individual’s exchanges with the social and material realities they face in virtue of occupying gendered subject positions. This clarifies why individuals may either conform with or deviate from the gender norms they sense are relevant to themselves vis-à-vis their gendered subject positions, because they can avow, repudiate, or remain neutral toward their gender-related affects and thoughts, working sometimes to cultivate them and other times to change them, and, more broadly, to cultivate or change the social positions which inculcate them. Accordingly, Jenkins defines ‘gender as identity’ such that:

(i) S has a gender identity of X iff S’s internal ‘map’ is formed to guide someone classed as a member of X gender through the social or material realities that are, in that context, characteristic of Xs as a class.

Thus, on Jenkins’s norm relevancy account, to have a feminine gender identity just means to take a preponderance[4] of feminine norms as relevant to oneself *because* of having formed an

[4] For terminological ease I use ‘preponderance’ here, but Jenkins’s more precise specification is that someone take a significant subset of one type of gender norms as relevant to themselves and not take a significant subset of another

embodied, internal map to guide someone classed as a woman through the social or material realities that are characteristic of women as a class in some context. Similarly, to have a masculine gender identity on the norm relevancy account just means to take a preponderance of masculine norms as relevant to oneself *because* of having formed an embodied, internal map to guide someone classed as a man through the social or material realities that are characteristic of men as a class in some context. Dissimilarly, on the norm relevancy account, to have a nonbinary gender identity (where 'nonbinary' refers to the class of nonbinary identities) is to take *no* feminine or masculine gender norms as relevant to oneself, although Jenkins points out her account can be used to give more specific nonbinary identities. For example, she defines the specific nonbinary identity 'genderfluid' as someone taking a preponderance of masculine or feminine (or both) gender norms as relevant to themselves at different times or in different contexts.

Having reviewed Jenkins's norm relevancy account, it is time to adapt it to do the work I set out to accomplish in the introduction, namely, analyzing the self-identification manifest by gender self-ascriptions. As we shall see, Jenkins's idea of taking a preponderance of gender norms as relevant to oneself vis-à-vis ones embodied, internal map of gendered subject positions will play a necessary but not sufficient role in how individuals self-identify.

3 THE NORM RELEVANCY ACCOUNT, THE SELF-IDENTIFICATION ACCOUNT, & PSYCHOLOGICAL GENDER

Consider an agender person. Let's call them 'Fate'. Prima facie, to be agender is not to self-ascribe any gender and to want to be perceived as a person without one (e.g., see Dembroff, 2020, pp. 9-10). Let's assume for the sake of argument, then, that our agender protagonist, Fate, does not self-ascribe a gender and wants to be perceived as a person without a gender. Now, let's

type of gender norms as relevant to themselves (Jenkins, 2018, p. 731). I take 'preponderance' to capture this idea rather well, but in cases where readers disagree, please defer to Jenkins's original formulation.

also assume that Fate was assigned female at birth and raised as a woman, but came to identify as agender later during college once they stopped caring about feminine or masculine norms, even though they still sometimes feel external pressure to conform to them.

How would the norm relevancy account classify Fate? Since they were assigned female at birth and raised as a woman, it stands to reason Fate would have formed an embodied, internal map to navigate the social and material realities characteristic of women as a class in their context. Moreover, in virtue of having formed such an embodied, internal map, it follows that Fate would take a preponderance of feminine norms as relevant to themselves vis-a-vis their gender-related affects and thoughts prompted while navigating that context. However, we saw that Fate stopped caring about feminine and masculine norms during college. Since on the norm relevancy account the sole criterion for having a feminine gender identity is to take a preponderance of feminine norms as relevant to oneself, the norm relevancy account would thus accurately classify Fate as not having a feminine gender identity. But Jenkins's account leaves the question of how Fate came to have their agender identity unanswered, since at one point they felt a predominance of feminine norms were relevant to them and thought about themselves one way, and now they feel and think of themselves differently. One way to put this is that as it stands, the norm relevancy account cannot explain the link between individual's predominant experiences of norm relevancy and their beliefs and desires about their gender[5]. Despite having been raised as a woman, Fate believes they are agender and desires to be seen as a person without a gender, likely in resistance to significant external pressures to identify and present in a feminine way.

[5] One might also think that people's beliefs and desires might not be sensitive to the norms they take as relevant to themselves because of phenomena like self-deception, which Talia Bettcher discusses (2009). One way to handle this issue is just to claim that despite various self-deceptive phenomena affecting individual's beliefs and desires about their gender, they will nevertheless have experiences of norm relevancy guiding them to be masculine or feminine in a context with which they can form more authentic beliefs and desires.

This shows us that to explain the relationship between individual's experiences of norm relevancy and their beliefs and desires about being a gender, we need to explain the missing link in the causal chain between them. I claim the missing link is the gender self-concept and gender role ideal individuals form from their experiences of norm relevancy, although the self-concept and role ideal can and often do come apart. On my 'self-identification account' of gender psychology, the gender self-concept consists of either a belief or contextual assumption representing individuals as a gender (Thagard, 2012; Thagard & Wood, 2015). Additionally, the gender role ideal consists of the expectations people have about what the role requires and their standards for performance (Zhang, 2018). Individual's gender self-concept enables them to appraise themselves relative to the expectations and standards of their gender role ideal, and that consequently enables them to desire to perceive themselves and be perceived by others as exemplifying it insofar as they avow their appraisals (Fausto-Sterling, 2021, p. 9). However, the gender self-concept also enables people to deliberately deconstruct and/or disavow their gender role given epistemic, pragmatic or political considerations, such as in cases where individuals adopt a gender-nonconforming ideal that aims to undermine the associated expectations and standards (Watson, 2015). Thus, I understand gender self-identification as consisting of the formation of a gender self-concept and *sometimes* of a gender role ideal on the basis of individual's experiences of norm relevancy, which will sometimes involve reflection on those experiences. Accordingly, I define gender self-identification such that:

S self-identifies as gender G if and only if:

(i) S has an embodied, internal map of their subject position that corresponds to G.

(ii) On the basis of that embodied, internal map, S either believes or assumes they are a G.

(iii) On the basis of their embodied, internal map and assumption or belief they are a G, S could desire to perceive themselves and be perceived by others as a G.

As we can see, the first component corresponds to Jenkins's embodied, internal map of gendered subject positions. This element is essential for explaining gender self-identification because the crucial factor determining what gender an individual self-identifies as will be what the available gender ideology is in their context at a given point in the lifespan (see Geuss, 1981, for an account of ideology). This is because gender ideology determines not only individual's experiences of norm relevancy vis-à-vis what gender-related affects and thoughts they have, but also their self-reflection about them insofar as it structures the concepts they use to form beliefs and desires about being a gender. According to Dembroff (2020), binary gender ideology is currently dominant around the world, and posits the two discrete and mutually exclusive gendered social categories 'man' and 'woman'. These gender categories have historically been associated with characteristic sets of innate physical (e.g., gonads, genitalia) and psychological (e.g., personality, intelligence) traits (see Hyde et al, 2020, for a critical discussion of brain sex dimorphism; and Mikkola, 2018, for a critical discussion of the distinction between sex and gender). As we shall see below, binary gender ideology plays a crucial role in how individuals form their beliefs about being a gender by structuring experiences of norm relevancy and concepts of gender categories.

This brings us to the second component in the definition, which corresponds to individual's belief or context-based assumption they are a gender. These components are important for explaining gender self-identification because without assuming or believing they are a gender kind, individuals would experience their gender-related affects and thoughts as foreign. Consequently, they would be unable to use their gender self-concept as a basis for appraising their gender role performance or altogether resisting their gender role using those affects and thoughts. Accordingly, I claim that to form a gender self-concept, individuals may either intuitively or self-reflectively infer their belief they are a gender kind through their experiences of norm relevancy. Both the intuitive and reflective inferences rely on a match between individual's gender concept(s) on the one hand and their gender-relevant physical and psychological traits, as indicated by their gender-related affects and thoughts, on the other (McKitrick, 2007; Fausto-Sterling, 2021). However, I leave it open that sometimes individuals will form a gender self-concept through assumption in a context, as when people questioning their gender "try on" different ways of identifying or presenting, which in any case still rely on the same gender concepts. According to Jennifer McKitrick, these family resemblance concepts of gender are based on "an indeterminate cluster of interrelated traits" like "primary and secondary sex characteristics, modes of dress and grooming, personality, preferences, occupations, expectations, and relationships", all of which allow individuals to "categorize [themselves] as masculine or feminine if [they] have enough of [these] characteristics, to a sufficient degree" (McKitrick, 2007, p. 144). Moreover, "in societies with exactly two well-defined gender norms" like our own, "individuals feel pressured to exemplify one cluster of characteristics, to the exclusion of the other" (McKitrick, 2007, p. 147). Finally, as McKitrick points out, "to identify with a group is to feel you are similar to members of that group and that you are or should be part of that group" (McKitrick, 2007, p. 139). Thus, as individuals have sufficiently many, sufficiently strong experiences of norm relevancy – understood as gender-related affects and thoughts – guiding and directing them to behave femininely or masculinely, they gradually learn to associate themselves with women and men because of them (Fausto-Sterling, 2021, pp. 7-10). This association between the sense of self and gender concepts can be understood as a linkage between what John Perry calls the 'self-notion', defined as "being the repository for information gained in normally self-informative ways and the motivator for actions done in normally self-effecting ways", and what he calls 'files' in memory, for example representing gender concepts like 'man' or 'woman' (Perry, 1998c, p. 325; 2001b1, p. 50; 2001b2, pp. 120-121; 2002c, p. 205; Van Leeuwen, 2012, p. 103)[6]. Establishing this linkage is how, to put it in Jenkins's terms, individuals come to viscerally feel a "sense of themselves as a man, woman, or some other gender" (Jenkins, 2018, 714). This is because after associating themselves with women or men on the basis of their gender-relevant physical or psychological traits, individuals come to think of themselves *as* women or men and often attune their behavior to the corresponding norms. Normally, this process of inferring gender happens early in children's development, with individual's belief in their status as a boy or girl cemented by age three (most likely via intuition given the limits of self-reflection at this age) (Diamond, 2020, p. 1). However, this belief is not immutable, and can change with sufficient changes in individual's experiences of norm relevancy over the lifespan, as exemplified by some trans and nonbinary people who begin reflectively spurred gender transitions during later childhood, adolescence, or adulthood (Berenbaum, 2018; Gülgöz and colleagues, 2019).

Finally, the third component in the definition corresponds to individual's desires to perceive themselves and be perceived by others as their gender. As we saw, these desires are based on individual's self-appraisal relative to their gender role ideal, which denotes individual's expectations about what the role consists of and standards for how to perform it. This feature is important for a definition of gender self-identification because individuals *often do* form desires

[6] Perry analyzes 'normally self-effecting ways of knowing' as a subclass of what he calls 'agent-relative knowledge', which relates agents to their own bodies without representing them (e.g., pain in one's leg). In contrast, he analzyes 'normally self-effecting ways of acting' as a subclass of action in which the agent is the only one affected by them (e.g., scratching one's itch). See Van Leeuwen, 2012, for more details about Perry's theory of the self-notion.

to conform to these expectations and standards (e.g., see Jones and colleagues, 2019), sometimes intentionally and sometimes not. Accordingly, I understand individual's 'gender role ideal' to be jointly constituted by their embodied, internal maps of gendered subject positions and their belief or assumption they are a gender kind. This feature of joint constitution is crucial because, as we saw, individual's beliefs or assumptions about their gender are what enable them to use their gender-related affects and thoughts to navigate their gendered subject positions. Thus, the expectations and standards individuals have about performing their gender role are expressed through the gender-related affects and thoughts they have while navigating a context. These gender-related affects and thoughts allow individuals to appraise their performance of their gender role and take steps to enhance it in context-sensitive ways. This is because they can be prompted by external stimuli, as when individuals are confronted with social situations that demand gender conformity (e.g., anxiously waiting to speak as a woman in a male dominated space), or internal stimuli, as when they reflect on their gender-relevant physical or psychological traits (e.g., excitedly imagining how one would look following facial feminization surgery). Insofar as individuals avow their appraisals, they subsequently desire to exemplify that ideal in their own and other's eyes in that context (Fausto-Sterling, 2021, p. 9). This clarifies why individuals become motivated to conform to the expectations and standards for their gender role in context-specific ways depending on whether a certain gender-relevant physical or psychological trait is expected in some, most, or all contexts (e.g., the feminine norm of having breasts applies across contexts, whereas the feminine norm of being demure applies more parochially). In virtue of representing themselves as a gender through their self-concept, people's self-esteem often becomes sensitive to their self-appraisals relative to their gender role ideal. Consequently, individuals often feel better or worse about themselves consistent with how well they perceive themselves or how well they think others perceive them as satisfying that gender role ideal, which motivates them to enhance their performance. All that said, as discussed briefly above, sometimes people decide to deconstruct and/or disavow their gender role ideal for political, epistemic, or pragmatic reasons. This requires some self-reflection to divorce one's gender status from one's self-assessment, such that an individual ceases to assess themselves according to how well they perform their gender role. It also requires some self-control to ensure that individuals do not engage in the habit of assessing themselves according to a previously held gender role ideal, at least until the habit is extinguished.

Now that we have discussed the rationale for my account of self-identification, let's turn to Dembroff's objection to the norm relevancy account (§3) and then consider how self-identification can shed light on cis, trans and nonbinary people's gender self-ascriptions (§4)[7]. As we shall see, the self-identification is the same for all three classes of identities. Although the details differ between them, in each case individuals reflect on their experiences of norm relevancy, form a belief about their gender, and then form desires to exemplify it relative to a role ideal.

4 DEMBROFF, THE SELF-IDENTIFICATION ACCOUNT, & PSYCHOLOGICAL GENDER

In their work on the norm relevancy account, Robin Dembroff (2020) argues that it cannot explain the nonbinary gender identities 'nonbinary woman' and 'nonbinary man'. Dembroff's argument for why the norm relevancy account cannot explain these gender identities hinges on the fact that they are *dual identities*. What it means to be a nonbinary woman or man, according to Dembroff, just *is* to identify as 'nonbinary' *and* as a man or woman (Dembroff, 2020, p. 9). As

[7] I use the terms 'cis', 'trans', and 'nonbinary' here to refer to the classes of individuals commonly denoted by these umbrella terms (Bettcher, 2009, section 1). I do not consider 'cis', 'trans' and 'nonbinary' versions of the gender identities of 'woman' and 'man' to be essentially different except insofar as they develop consistently or inconsistently with the sex assigned at birth.

they point out, when we consider how the norm relevancy account would classify individuals with these dual identities, the account yields a contradiction. On the one hand, the norm relevancy account requires that someone take no feminine or masculine norms as relevant to themselves to have a 'nonbinary' gender identity. On the other hand, the norm relevancy account requires that someone take a preponderance of feminine or masculine norms as relevant to themselves to have the gender identity of 'man' or 'woman'. Consequently, to identify as a nonbinary woman or nonbinary man on the norm relevancy account would require that someone both take and not take a preponderance of feminine or masculine norms as relevant to themselves, which is a contradiction. Ergo, Dembroff concludes the norm relevancy account cannot explain these gender identities. But while I think Dembroff is right to conclude the norm relevancy account succumbs to this explanatory problem, I will show in the next section that my account of self-identification does not. I will do this by considering a series of cases illustrating how cis, trans, and nonbinary individuals self-ascribe the gender identities 'woman', 'nonbinary woman', and the non-gender identity 'agender', while being careful to point out what would differ in the case of their counterpart gender self-ascriptions (i.e., 'man' and 'nonbinary man'). My goal in performing this exercise is to showcase first that my account of self-identification can systematically explain the gender identities of interest, namely cis, trans, and nonbinary ones, and second that an account of gender identity with a norm relevancy component is able to explain them better than an account without it. For easy reference, I will use the terms 'map', 'belief', and 'desired perception' in bold to signal when the respective clauses in my definition of self-identification are being invoked in each case. Additionally, I will use the term 'sub-map' to refer to different sets of norms that are represented in one's overall map.

5 SELF-IDENTIFICATION & CIS, TRANS, AND NONBINARY GENDER IDENTITIES

Consider a hypothetical cis woman named ‘Zoe’. Assume Zoe was assigned female at birth, raised as a girl in a context where binary gender ideology was dominant, and now self-ascribes the gender identity ‘woman’ as an adult. Assume also that Zoe has loved her female-coded body and feminine presentation since she was a young girl. How can the self-identification account of gender identity explain Zoe’s gender self-ascription? First, since Zoe was assigned female at birth, raised as a girl, and loved feminine presentation, she would have developed a **map** of the subject position ‘woman’ and would have correspondingly predominant experiences of norm relevancy for femininity. Second, on the basis of having sufficiently many, sufficiently strong varieties of these experiences, Zoe would have formed a **belief** she was a girl. Third and finally, on the basis of her **map** of the subject position ‘woman’ and her **belief** she is a girl, Zoe would have **desired perceptions** for feminine presentation. Thus, Zoe self-ascribes the gender identity ‘woman’ because her **map** of that subject position, her **belief** she is a woman, and her **desired perceptions** for femininity constitute her self-identification as a woman. *Mutatis mutandis* for a hypothetical cis man, as if we had been considering a case involving one, the only differences would be that the **map** would be of the subject position ‘man’, the **belief** would be that they are a man, and the **desired perceptions** would be for masculine presentation.

Now that we have seen how my account of self-identification explains cis women’s and cis men’s gender self-ascriptions, let’s turn to a case involving a hypothetical trans woman named ‘Alicia’. Assume Alicia was assigned male at birth, raised as a boy until twelve in a context where binary gender ideology was dominant, and now self-ascribes the gender identity ‘woman’ as an adolescent. Also assume that Alicia felt neutral about her male-coded body and loved feminine presentation from the time she was a young child. How would the self-identification account explain Alicia's gender self-ascription? First, since Alicia was raised as a boy but loved feminine presentation, she would have formed a **map** of the subject position 'woman' and would have corresponding experiences of norm relevancy for femininity. Second, on the basis of having sufficiently many, sufficiently strong experiences of norm relevancy for femininity, Alicia would have formed a **belief** she was a girl. Third and finally, on the basis of her **map** of the subject position 'woman' and her **belief** she is a girl, Alicia would have **desired perceptions** for feminine presentation. Thus, Alicia self-ascribes the gender identity 'woman' because her **map** of that subject position, her **belief** she is a woman, and her **desired perceptions** for femininity constitute her self-identification as a woman. *Mutatis mutandis* for a hypothetical trans man, as if we had been considering a case involving one, the only differences would be that the **map** would be of the subject position 'man', the **belief** would be that they are a man, and the **desired perceptions** would be for masculine presentation.

Since we have discussed how my account of self-identification explains trans women's and trans men's gender self-ascriptions, let's turn to a case involving a hypothetical nonbinary woman named 'Destiny'. Assume Destiny was assigned female at birth, raised as a girl in a context where binary gender ideology was dominant, and now self-ascribes the gender identity 'nonbinary woman' as an adult. Assume also that Destiny has loved her female-coded body and masculine, "tom-boy" presentation since she was a young girl. Finally, assume that after learning about nonbinary identities in a feminist philosophy course in college, Destiny came to repudiate the binary gender ideology holding that sex and gender are identical, that gender is necessarily binary, and that femininity and masculinity are mutually exclusive. How can the self-identification account of gender identity explain Destiny's gender self-ascription? To answer this question, we must consider two self-identifications: first, Destiny's self-identification as a woman, and second, her self-identification as nonbinary. First, since Destiny was assigned female at birth, raised as a girl, and loved feminine presentation, she would have developed a **map** of the subject position 'woman' and would have correspondingly predominant experiences of norm relevancy for femininity. Second, on the basis of having sufficiently many, sufficiently strong varieties of these experiences, Destiny would have formed a **belief** she was a woman. Third and finally, on the basis of her **map** of the subject position 'woman' and her **belief** she is a girl, Destiny would have **desired perceptions** for feminine presentation. Thus, Destiny would initially self-ascribe the gender identity 'woman' because her **map** of that subject position, her **belief** she is a woman, and her **desired perceptions** for femininity constitute her self-identification as a woman. Now that we have seen how Destiny's initial self-identification as a woman comes about, let us turn to her self-identification as a nonbinary woman (here I will use 'sub-map' to refer to different sets of norms that one's map tracks, but this is only a convenient rhetorical tool and is not meant to imply that people have *multiple* such maps). First, since Destiny took a feminist philosophy course, learned about nonbinary identities, and came to believe that sex and gender are distinct, that gender is not necessarily binary, and that femininity and masculinity are not mutually exclusive, she came to develop a **sub-map** of the subject position 'nonbinary' as being one that flouts binary gender ideology. Second, through self-reflection about her recently formed metaphysical beliefs about gender and her experiences of norm relevancy for both the feminine and masculine subject positions, she begins to feel less limited by her initial feminine **sub-map** and so forms a **belief** that she is nonbinary. Third, on the basis of her **sub-map** of the subject position 'nonbinary' and her **belief** she is nonbinary, Destiny forms **desired perceptions** for nonbinary presentation[8], which

[8] By 'nonbinary presentation', I just mean gender presentation that deviates from binary gender norms. I do not mean to imply that there should be *any* standards for nonbinary presentation.

given her metaphysical beliefs about gender and her mixed experiences of norm relevancy for feminine body and masculine dress appearance amount (in her case) to a desire for an androgynous presentation. Thus, Destiny would secondarily self-ascribe the gender identity ‘nonbinary’ because her **sub-map** of that subject position, her **belief** she is nonbinary, and her **desired perceptions** for nonbinary presentation constitute her self-identification as nonbinary. Now that we have discussed Destiny’s self-identification as nonbinary, it is easy to see how she self-ascribes the gender identity ‘nonbinary woman’. On the basis of her two **sub-maps** of the subject position’s ‘woman’ and ‘nonbinary’, her **beliefs** she is a woman and nonbinary, and her **desired perceptions** for feminine and nonbinary presentation, Destiny self-identifies as a nonbinary woman. *Mutatis mutandis* for a hypothetical nonbinary man, as if we had been considering a case involving one, the only differences would be that the **sub-maps** would be of the subject position’s ‘man’ and ‘nonbinary’, the **beliefs** would be that they are a man and nonbinary, and the **desired perceptions** would be for masculine and nonbinary presentation.

Having covered how my account of self-identification explains the gender identities ‘nonbinary woman’ and ‘nonbinary man’, let’s close by returning to the case involving our agender protagonist ‘Fate’ to see how it explains ‘agender’ identities. Recall that Fate was assigned female at birth and raised as a girl in a context where binary gender ideology was dominant, but later became agender after finding they no longer cared about gender norms. How would the self-identification account explain Fate’s agender identity? To see how, we need to consider two self-identifications: first, Fate’s self-identification as a woman, and second, their self-identification as agender. First, since Fate was assigned female at birth and raised as a girl, they would have developed a **sub-map** of the subject position ‘woman’ and would have correspondingly predominant experiences of norm relevancy for femininity. Second, on the basis of having sufficiently many, sufficiently strong varieties of these experiences, Fate would have formed a **belief** they were a girl. Third and finally, on the basis of their **sub-map** of the subject position 'woman' and their **belief** they are a girl, Fate would have **desired perceptions** for feminine presentation. Thus, Fate initially self-ascribed the gender identity 'woman' because their **sub-map** of that subject position, their **belief** they are a woman, and their **desired perceptions** for femininity constitute their self-identification as a woman. Now that we have seen how Fate's initial self-identification as a woman came about, let us turn to their self-identification as agender. Ever since Fate stopped caring about gender norms, they ceased to **believe** that they are a woman because of being unmoved by their experiences of norm relevancy for femininity. Moreover, since Fate ceased to **believe** they are a woman, they ceased to have **desired perceptions** for feminine presentation on the basis of their belief *despite* the fact that they still have experiences of norm relevancy for femininity. Since Fate does not satisfy clauses (ii) and (iii) of the definition of self-identification, they just *lack* a gender identity of any sort, which clarifies why, unlike Zoe and Destiny, they lack any gender self-ascription whatsoever.

Now that we have discussed how my account of self-identification can explain the various gender identities of interest, it is time to turn to a discussion of how my account of self-identification promises to help the neuroscience of gender.

6 SELF-IDENTIFICATION, GENDER IDENTITY, & CONSTRUCT STABILIZATION

The neuroscience of gender identity has been developing for the past three decades (Caselles, 2021; Gonzalves, 2020), and today there are two theories of gender identity currently competing in the neuroscience literature: what I will call the 'cortical development and self-attribution hypothesis' offered by Uribe and colleagues (2020), and what I will call the 'cognitive development and gender socialization hypothesis' published by Anne Fausto-Sterling (2021). Both are causal theories insofar as they purport to explain what *causes* gender identity to develop. But they are undermined both by plausibly inaccurate participant assignment in experimental studies and their vague specification of the phenomenon, since they lack conceptual definitions of gender identity that can be used to capture individual differences between trans and nonbinary participants and their existing definitions do not offer a substantive constitutive theory specifying what gender identity *is*. Indeed, this lack of clarity about participant assignment and what constitutes gender identity obscures whether these causal theories have identified instances of cognitive capacities involved in gender identity or discovered how the brain supports them (Sullivan, 2012). I claim these theory's explanatory shortfalls are due both to a lack of intradisciplinary collaboration between neuroscientists, which has played a role in their proposing conceptual definitions of gender identity that cannot partition groups accurately, and to a lack of interdisciplinary collaboration between neuroscientists and psychologists, which has played a role in the former's lack of a constitutive theory of gender identity. In what follows, I will first explain some notions from the philosophy of science concerning when a theory of a phenomenon matches phenomena observed in the lab and how interdisciplinary collaboration in theory development plays a role in ensuring this match is obtained (§5.1). Afterward, I will explain why a lack of interdisciplinary collaboration has prevented neuroscientists theorizing about gender identity from attaining this match (§5.2). Finally, I will offer an antidote to this theoretical nadir by showing how my constitutive theory of gender identity as a kind of 'self-identification' can help neuroscientists secure this match (§5.3).

6.1 Construct Stabilization in the Mind-Brain Sciences

According to Jacqueline Sullivan, matching what cognitive capacities and or neural components and processes neuroscientific theory claims constitute a phenomenon to observations of cognitive capacities or neural activity in the lab is best achieved through a kind of interdisciplinary theory development. To see why this is the case, we need to briefly review what she calls ‘construct validity’, which pertains to the match, and what she calls ‘construct stabilization’, which pertains to the interdisciplinary theory development she argues helps to secure this match. But first, some preliminary remarks on neuroscientific explanations, constructs, and experimental paradigms will prove helpful in understanding construct validity and construct stabilization as Sullivan construes them.

Following the influential mechanist account of explanation in neuroscience, Sullivan says neuroscientists offer *mechanistic* explanations[9]. These integrate explanations of the functional role of cognitive capacities and their component subcapacities traditionally offered by cognitive psychologists with explanations of neural components and processes at various levels of neural composition traditionally offered by cognitive neurobiologists. To produce mechanistic explanations, *cognitive neuroscientists* use the behavioral tasks developed by *cognitive psychologists* to evoke target cognitive capacities and their component subcapacities alongside neuroscientific techniques for neural imaging, recording, and intervention to functionally localize and attribute neural components and processes to them. A complete, multilevel mechanistic explanation thus specifies what neural components and processes realize specific cognitive capacities and subcapacities at different levels of neural organization.

[9] For an introduction to mechanist philosophy of neuroscience, see Machamer, Darden, and Craver (1999). For criticisms of the scope of the mechanist account of explanations in neuroscience, see Ross (2018) and Bickle (2020).

Constructs, then, are concepts that group instances of cognitive capacities, and Sullivan thinks neuroscientists use constructs to create experimental paradigms to produce, measure, and detect these cognitive capacities and their associated neural machinery. These experimental paradigms specify what stimuli are to be presented and how, what the response variables to be measured are across the phases of the experiment and how to measure them, and what the comparative measurements of these response variables must equal to ascribe a cognitive capacity to an organism and/or the location of a function to a given brain area.

Thus, for Sullivan, construct *validity* has to do with whether the experimental paradigms cognitive psychologists and cognitive neuroscientists use to study constructs produce, measure, and detect instances of the cognitive capacities and component subcapacities they denote. She thinks cognitive psychologists and cognitive neuroscientists alike ascertain the degree of construct stability using what she calls ‘construct explication and assessment’, which she also thinks plays a role in how these scientists stabilize constructs using what she calls ‘perspectival pluralism’. On the one hand, construct explication and assessment consist of a series of questions including (1) which phenomena will be grouped under the concept designating a construct, (2) what investigative strategies will produce instances of it, (3) whether the investigative strategies are adequate or should be modified, and (4) what available data entail about whether the construct should be revised to include or exclude certain phenomena. On the other hand, perspectival pluralism corresponds to the different ontological perspectives that cognitive psychologists and cognitive neuroscientists use to study cognitive systems insofar as they invoke different “variables to characterize and partition those systems into parts” like cognitive capacities and subcapacities or neural components and processes (Sullivan, 2012, p. 671).

Crucially, Sullivan thinks that to achieve *construct validity*, researchers must stabilize constructs by performing construct explication and assessment in a way that incorporates perspectival pluralism. In other words, Sullivan thinks construct explication and assessment should be performed through interdisciplinary collaboration to include the perspectives of cognitive neuroscientists and cognitive psychologists alike. For Sullivan, this is the only way to achieve what she calls *construct stability*, which consists of agreement between scientists working in different labs situated in the same and different areas of science on how to define the terms for constructs, what the best experimental paradigms are to study a construct, and when two experimental paradigms may be said to measure the same cognitive capacity denoted by a construct. Sullivan thinks this because, as we saw, providing mechanistic explanations requires both identifying cognitive capacities and their component subcapacities using behavioral tasks drawn from cognitive psychology *and* localizing or attributing their functions to neural components and processes at different levels of neural composition using imaging, recording, and intervention techniques taken from different areas of neuroscience. This means that there will be a plethora of potential variables and ways of using them to partition cognitive systems that may be relevant for individuating a target cognitive capacity or subcapacity and associating it with different brain regions and processes. In turn, this entails researchers from different areas of science will be able to stabilize constructs *faster* if they collaboratively design constructs and experimental paradigms to test them, since this will allow them to come to agreement more quickly about the terms, methods, and measures to be used.

To close out this review of construct validity and stability, let’s briefly consider the example Sullivan gives of how construct explication and assessment incorporating perspectival pluralism helped stabilize the construct of place learning. In his original study of this construct, the cognitive neuroscientist Richard Morris defined it as "the cognitive ability to find a hidden target in the absence of local cues" and adopted an "information processing view of the mind" which shaped his design of the so-called water maze, the experimental paradigm with which he tried to isolate this cognitive capacity and localize it to a brain area (Sullivan, 2012, p. 671). Morris's original experiments led him to conclude the water maze individuated the cognitive capacity of place learning and localized it to the rat hippocampus. But later experiments in which Morris and colleagues adopted "an information processing view of the brain and its structures" showed rats with hippocampal lesions could still navigate the maze successfully, indicating that "the water maze does not individuate a discrete cognitive capacity" and that instead "[associative and nonspatial] cognitive capacities are [also] involved" (Eichenbaum, Stewart, and Morris, 1990; Sullivan, 2012, p. 671). In contrast to Morris and colleague's use of the water maze to individuate these component cognitive subcapacities in place learning and localize them to the hippocampus, cognitive neurobiologists were using the water maze to study cellular and molecular activity in the hippocampus *without* adopting this information processing view of the brain. This made cognitive neurobiologists insensitive to the fact that the water maze does not individuate a single cognitive capacity, which contributed to the instability of the construct of place learning insofar as it led them to posit mechanistic explanations "without clear explananda, like the claim that NMDA-receptor activation in the hippocampus is a necessary component of the mechanism of *spatial memory*" (Sullivan, 2012, p. 671, emphasis added). In response, Morris enlisted the help of researchers trained in cognitive psychology to ascertain why rats with blocked NMDA-receptors fail to navigate the water maze. They used "a battery of cognitive tests designed to identify what informational processes are disrupted by NMDA-receptor blockade and what information rats . . . learn in the water maze" (Sullivan, 2012, p. 671). Having taken this information processing perspective of the brain and incorporated these cognitive tests, Morris's interdisciplinary team was able to demonstrate that NMDA-receptor activation likely "disrupts non-spatial as well as spatial components of water maze learning" (Bannerman and colleagues, 1985, p. 185). Thus, by performing construct explication and assessment in a way that incorporated perspectival pluralism, Morris and colleagues were able to design an experimental paradigm that helped stabilize the construct of place learning. This is because incorporating cognitive tests allowed Morris and colleagues to show that contrary to what cognitive neurobiologists claimed and consistent with their prior research, NMDA-receptor activation impedes both *spatial* and *non-spatial* cognitive processes involved when rats navigate the water maze.

6.2 Assessing the Construct of Gender Identity in Neuroscience

Now that we have discussed the relevant background on construct stability and validity, it is time to apply these notions to assess the status of the neuroscientific construct of gender identity. Accordingly, I will argue the lack of collaboration between neuroscientists working in different areas of neuroscience and between neuroscientists and psychologists has resulted in the joint instability and invalidity of the construct. The lack of intradisciplinary collaboration between neuroscientists and their resulting disagreement about terms has contributed to the design of insufficiently precise conceptual definitions of gender identity, which undermines the construct's validity since the definitions cannot be used to partition study groups to capture individual differences in the phenomenon. Furthermore, the lack of interdisciplinary collaboration between neuroscientists and psychologists and their consequent disagreement about terms for the construct has contributed to neuroscientist's lacking a constitutive theory of gender identity. This second source of construct instability is the primary contributor to the invalidity of the construct, as it obscures whether neuroscientists have identified any cognitive capacities involved in gender identity and what experimental paradigms should be used to study them. To see why these problems arise, we need to briefly state these theory's main claims about what gender identity is and how it develops, and then discuss the disagreements between neuroscientists and between neuroscientists and psychologists on what terms should be used to study gender identity.

The two neuroscientific theories of gender identity we will discuss are what I call the 'cortical development and self-attribution hypothesis' and the 'cognitive development and gender socialization hypothesis'. According to the cortical development and self-attribution hypothesis, which is posited by Uribe and colleagues (2020), gender identity is a 'feeling of belonging to the male or female sex', and its development is determined by the sexual differentiation of neural networks involved in identifying the body as part of the self. This theory claims the sexual differentiation of the brain is achieved by a developmentally timed process of cortical thinning, understood as a combination of synaptogenesis and pruning[10], that produces sex dimorphic patterns of cortical thickness (i.e., gray matter volume) in various brain areas in response to dominant levels of androgens or estrogens[11]. Moreover, it claims that cortical thinning develops the four interacting networks that allow individuals to make bodily self-attributions, including the executive network, associated with deliberate attention, self-control, and planning, the default mode network, associated with imagination and mind wandering, the salience network, associated with automatic attention and the integration of various sensory modalities (e.g., sight, touch,

[10] Synpatogenesis is the process by which new synapses are formed, where a synapse is understood as the minute gap between a presynaptic neuron's axons and a post-synaptic neuron's dendrites which chemical transmitters diffuse across.

[11] Note that this claim is undermined by the fact that the construct of brain sex is itself unstable, as neuroscientific researcher Daphna Joel disagrees with Uribe and colleagues about what experimental paradigms should be used to study brain sex. Joel's lab (Joel, 2021) has consistently shown the brains of male-designated and female-designated participants show *mosaic* rather than *dimorphic* patterns of cortical thickness, which she claims undermines the predictive and explanatory power of the construct of brain sex. It is also undermined by the fact that structure and function is determined by the interplay between genes and the environmental stimuli that activate them rather than solely by genes (Jordan-Young, 2010).

audition, proprioception, etcetera), and the sensorimotor network, associated with sensory perception and motor control. The central claim of the theory is thus that the executive network allows individuals to toggle selective attention between default mode network representations of the self and salience- and sensorimotor-network driven representations of body- and social perceptions to build up a gendered self-image. Since the original cortical development hypothesis held that the sexual differentiation of the brain creates structural and functional differences between men and women's brains that create corresponding differences in their behavior, personality, and feelings of belonging to the group of men or women, it stands to reason that Uribe et al think self-representations of gender identity are built up through body-identification in a way that is ultimately determined by sex, although they do not spell this out (Caselles, 2021).

In contrast, the cognitive development and gender socialization hypothesis posited by Anne Fausto-Sterling (2021) defines gender identity as 'gender/sex identity' and holds its development is determined by the ways in which cognitive development and gender socialization interactively entrain the capacities to recognize gender traits in and subsequently attribute gender category membership to oneself, others, or objects. On this theory, gender/sex identity development occurs in three stages across the first thirty-six months of a child's life, and progression through each stage is driven by interactions between children and social others who treat them in gender-specific ways according to the gender/sex they are perceived as. In the first stage spanning the first 12 months of life, infant-parent interactions simultaneously expose children's developing nervous systems to multisensory cues that form the associations underlying their gender schemas and inculcate desires to perform gender-appropriate behaviors, as parents scaffold children's performance of these behaviors by physically manipulating them (e.g., putting a glove in a boy's hand, brushing a girl's hair) and vocally labeling them with valenced and/or identity-conferring gender terms (e.g., ‘good boy’, ‘pretty girl’, etc.). Next, in the second stage spanning from twelve to twenty-four months, children’s acquisition of the ability for language and conceptual thought allows them to translate their preconceptual, association-based gender schemas into fully-fledged gender concepts or ‘generative models’ of gender/sex. With these generative models, children can deliberately imagine and experiment with gender/sex elements such as gender symbols, practices, identities, norms, etcetera, and ascribe these elements of gender to themselves, others, and objects. Finally, in stage three spanning twenty-four to thirty-six months, children’s gender/sex identity becomes increasingly stabilized through their self-socialization to norms for the gender/sex they self-categorized as in the second phase. This process of self-socialization is made possible by children’s acquisition of the ability for self-directed movement during the third phase, their abilities for language and conceptual thought acquired during the second phase, and their repertoire of gender-stereotyped behaviors and behavioral desires acquired in the first phase.

Now that we have clarified the main claims about what gender identity is and how it develops on these theories, it is time to discuss the disagreements between neuroscientists, and between neuroscientists and psychologists, on the terms to use to study gender identity. We can already see that there is disagreement between Uribe and colleagues and Anne Fausto-Sterling on how to define gender identity, making the construct unstable due to disagreement between researchers working in different areas of neuroscience. Whereas Uribe and colleagues offer a brief, binary definition invoking feelings of belonging to the groups of men or women, Fausto-Sterling offers a neutral description invoking the relationship between sex and gender. There is a tacit disagreement here between these researchers on the terms for the construct, since the former’s definition excludes the possibility of nonbinary gender identities, whereas the latter’s description is explicitly formulated to include any pattern of gender/sex development. This disagreement is likely due to their lack of collaboration on construct explication and assessment, and it renders Uribe and colleague’s construct invalid for omitting nonbinary gender identities as a class, since they are part of the phenomenon but not accounted for by their construct. Nevertheless, neither Uribe and colleague’s definition nor Fausto-Sterling’s description gives us a way to delineate *specific* nonbinary gender identities., which makes both their constructs invalid. Since Uribe and colleague’s criterion of ‘feeling belonging to the group of men or women’ excludes nonbinary identities and Fausto-Sterling does not offer a criterion in the first place, both conceptual definitions fail to demarcate different identities and omit relevant complexity. Moreover, they plausibly obscure relevant complexity through inaccurate participant assignment, as in cases where nonbinary participants are coded as trans men or women. We know from anonymous online population surveys like that of Jones and colleagues (2019) that nonbinary youth misrepresent themselves as trans men or women to participate in experimental studies for gender affirming care, and moreover that trans and nonbinary individuals experience gender dysphoria differently and for different reasons. This means there is a nontrivial chance neuroscientific data adduced for *both* these theories has been confounded by inaccurate participant assignment, since nonbinary participants coded as trans would skew the data given their different experiences with gender identity. Relatedly, neither Uribe and colleague’s definition nor Fausto-Sterling’s description offers us a constitutive theory of what gender identity is supposed to be that explains in sufficient detail what cognitive capacities and subcapacities are involved and why, which also undermines the validity of their constructs. Although Uribe and colleague’s tell us that the capacity for body identification is involved and tie it to subcapacities realized by different networks, they do not explain how body identification will be affected by sex, and their claim that the subnetworks that realize it are sex dimorphic is at odds with Manzouri and Savic’s (2018) claim that they are not.

Uribe and colleague’s also fail to respond to a plausible objection that the self-representations in the default mode network they claim underly individual’s gender identities are a product of decision making and socialization, since these representations vary between cultures and men and women themselves (Hyde and colleague’s, 2018). Finally, they do not consider that their own construct of brain sex is itself unstable. Studies by neuroscientists like Daphna Joel (2021) show evidence that human brains have mosaic rather than dimorphic patterns of cortical thickness, which undermines the claim that cis and trans people have the gender identities they do because of innate sex differences in the brain that cause downstream differences in behavior, personality, etcetera, making it unclear why body identification would yield different gender identities on their account. Moving on to Fausto-Sterling’s theory, although she strongly hints that the capacities for perception, personality, and self-categorization are involved in gender identity, she does not tell us how these capacities are related to neural networks besides the sensorimotor network, although she does much better at explaining how they are impacted by gender socialization. She also omits the role of theory of mind, which appears especially central to trans people’s gender identity development insofar as it is sometimes necessary for us to engage in the pretense of having a gender we disavow when we recognize others perceive us as a different gender than we identify as.

We saw from Sullivan that a good construct should specify what *specific* cognitive capacities and subcapacities are constitutive of a phenomenon at the outset so that experimental paradigms can be designed to identify and associate them with brain areas and processes. Spending too much time on construct design before conducting experiments can of course be a waste of time, as initial designs may give way quickly with the influx of data. Nevertheless, *some* careful thought should go into construct design to ensure that the phenomenon under study is not so misdescribed that it undermines getting any data that will advance theorizing. In this minimal sense of construct design, designing experimental paradigms before adequately designing the constructs designating the cognitive capacities and subcapacities they will be used to study puts the cart before the horse insofar as there is no guarantee the experimental paradigms will evoke instances of the *right* cognitive capacities. We also cannot ascertain whether existing experimental paradigms have evoked the right cognitive capacities without first specifying what the relevant cognitive capacities are in a way that spells out their function, how it is involved in a particular set of behaviors, and how those behaviors relate to environmental stimuli. This latter issue is especially important in the case of gender phenomena because it makes a difference whether the capacities are *defined* such that they necessarily only produce one of two gender identities, as Uribe and colleague's claim, or whether the capacities will be defined more neutrally such that they are just recruited by engagement in gender practices in the ways Fausto-Sterling claims. As we saw above, these neuroscientists haven't offered adequate definitional criteria to delineate existing gender identities nor a constitutive theory of gender identity that clearly specifies what cognitive capacities and subcapacities are involved and why in sufficient detail. Most dire, I think, were the issues that Fausto-Sterling does not explain what neural processes are recruited for the capacities she invokes, and Uribe and colleague's don't explain how gender socialization impacts body identification. Given these problems, it seems reasonable to conclude these researchers cannot be confident they have identified instances of the cognitive capacities and subcapacities denoted by their constructs, successfully localized or attributed them to brain areas, or identified experimental paradigms to study them. This all contributes to the invalidity of the construct of gender identity in neuroscience.

This case of construct invalidity is indicative of another disagreement between neuroscientists and psychologists on the terms for the construct, as cognitive psychologists have defined gender identity as both a personality trait and self-categorization to study how gender affects individual and group behavior (Wood and Eagly, 2015). Notice that the definitions of gender identity offered by these cognitive psychologists invoke specific cognitive capacities that are implicated within a social psychological account of gender role socialization, which explains how they are recruited to facilitate gender identity development. This constitutes a disagreement between neuroscientists and psychologists on the terms for the construct since each gives different definitions. Because psychologists like Wood and Eagly have used their definitions of gender identity fruitfully for several decades (see, e.g., Wood & Eagly, 2012), and since they are obviously relevant insofar as both neuroscientist's causal theories invoke personality traits and self-categorization, it stands to reason that there has been a lack of interdisciplinary collaboration between neuroscientists and cognitive psychologists on construct explication and assessment for gender identity. Crucially, this lack of collaboration and the resulting construct instability their disagreement on definitions entails seems principally to blame for why the neuroscientific construct of gender identity is invalid. As we saw, cognitive psychologists have special expertise when it comes to designing constructs and experimental paradigms that individuate specific cognitive capacities and subcapacities given their information processing view of the mind and repertoire of capacity individuating tasks. Although neuroscientists will have privileged information necessary to individuate and relate these capacities and subcapacities to neural components and processes insofar as they are tied to specific neural mechanisms, they can still benefit from psychologist's greater expertise in designing behavioral tasks that will play a role in this individuating process, and moreover they will need a robust psychological paradigm in the first place. Moving on, we also just saw that cognitive psychologists have developed definitions of gender identity that more finely specify the relevant capacities and subcapacities involved and how they are implicated in gender socialization, which could be used to design experimental paradigms in neuroscience. Thus, the facts that neuroscientists lack an adequate constitutive theory of gender identity and that their construct is consequently invalid seems best explained by the fact that they have not engaged in the kind of interdisciplinary theory development that Sullivan argues is necessary to stabilize constructs.

6.3 Gender Identity & Self-Identification

Now that we have discussed why the neuroscientific construct of gender identity is jointly unstable and invalid, it is finally time to discuss how my account of self-identification can help neuroscientists stabilize it to improve its validity. To see how my theory can do this work, we need to discuss how it can help neuroscientists and psychologists come to agreement about the definition of gender identity, the best experimental paradigms to study it, and when two experimental paradigms measure the same capacities associated with it. As we go, we will see that the tools my theory provides to help researchers stabilize the construct also offer them remedies to the problems undermining its validity.

Beginning with Sullivan's first criterion for construct stability, my theory of self-identification can help neuroscientists and psychologists come to agreement about the definitions of terms for the construct of gender identity. More specifically, it can help neuroscientists and psychologists come to agreement about the conceptual definition of gender identity by motivating them to cooperatively discuss or adopt my conceptual definition of gender self-identification, which can resolve the problem their own conceptual definitions face delineating gender identities. As we saw in sections two and four, my conceptual definition of gender self-identification offers a set of psychological components including Katherine Jenkins's embodied, internal map of gendered subject positions, beliefs and assumptions about one's gender kind status, and desires to perceive oneself and be perceived by others as that kind, which constitute individuals *gender self-concept* and *gender role-ideal*, respectively. These components provide fine-grained definitional criteria which can be used to partition cohort groups in experimental studies, since we also saw they track meaningful differences within and between groups with different gender identities. First, cis, trans, and nonbinary people each have different maps of gendered subject positions because they are socialized to different sets of norms (e.g., exclusively feminine norms for cis women, first masculine and then feminine norms for trans women, and what we might call 'feminist norms' for nonbinary women). This translates to these groups having detectable differences in their group attitudes and behaviors that can be used to partition them into participant groups. Moreover, differences along this axis can be tracked *within* and *between* these groups because my definition can delineate specific identities in each, enabling researchers to study differences between individuals holding different gender identities. This leads us to the second point, which is that cis, trans, and nonbinary people have different beliefs about gender, because the contents of these beliefs differ consistently with the differences in the gender concepts which inform them. For example, although trans men and women believe they are men and women just like cis people do, their gender concepts of 'man' and 'woman' are often different from cis people's insofar as they associate different traits with men and women (e.g., trans people have resistant notions like 'boy pussy' and 'girl dick' that explicitly associate the physical traits 'penis' and 'vulva' with both men and women). Furthermore, whereas cis people have latent, implicit or explicit beliefs that they are cis over and above their beliefs they are men or women, trans people who are aware of their trans status have explicit beliefs that they are trans over and above their beliefs they are men or women.

This translates into these groups having measurable differences in their beliefs that can also be used to partition them into participant groups, and as before, differences along this axis can be tracked *within* and *between* groups. Finally, cis, trans, and nonbinary people have different desired perceptions, because desired perceptions are informed by individual's beliefs about their gender kind status and map of their gendered subject position. For example, a conservative cis woman has a belief that she is a woman just as a nonbinary woman does, but the two have different maps with different group attitudes and behaviors with which they construe their role-ideals and appraise their performance. Whereas the conservative cis woman construes her role ideal and appraises her performance in accordance with dominant gender ideology, the nonbinary woman construes her role ideal and appraises her performance according to resistant gender ideology (e.g., from feminist philosophy). This translates to these groups having observable differences in their motivations which can be used to partition them into participant groups and to study differences in motivation within and between such groups. My definition thus provides three concrete operational criteria for partitioning study groups, allowing for the flexibility needed to study differences within and between groups[12], which addresses the problem I raised for the validity of the neuroscientific construct vis-à-vis accurately partitioning study groups.

This brings us to the next two criteria for construct stability offered by Sullivan, namely agreement between researchers in different areas of science on the best experimental paradigms to study a construct and when multiple experimental paradigms measure the same capacity denoted by a construct. My theory of gender self-identification can help neuroscientists and psychologists come to agreement on these questions by helping to specify what cognitive capacities are involved

[12] This has been sorely lacking in the neuroscience of gender identity due to the operationalization of neuroscientist's existing conceptual definitions of gender identity to *only* trans identities (Caselles, 2021)**.**

in gender identity. I claim the cognitive capacities involved in gender identity are whichever cognitive capacities are required to explain how people form and revise embodied, internal maps of gendered subject positions, beliefs about their gender kind status, and desired perceptions for it. A putative shortlist of cognitive capacities necessary for these components might include sensory perception, emotion, categorization, memory, theory of mind, attention, and metacognition. By pointing to these cognitive capacities, my theory provides neuroscientists and psychologists with the knowledge they need to cooperatively decide which experimental paradigms will target the many instances in which they are involved in gender identity and the many forms that involvement takes. Furthermore, by providing them with the knowledge they need to cooperatively identify the best experimental paradigms, my theory of gender self-identification also provides neuroscientists and psychologists with the knowledge they need to cooperatively come to agreement about when multiple experimental paradigms measure the same cognitive capacity. This is because they will be able to cooperatively evaluate the efficacy of experimental paradigms for measuring specific instances of these cognitive capacities given their knowledge that they are the relevant ones to study. My theory of gender self-identification thus helps specify what the relevant cognitive capacities should be for the construct of gender identity and provides direction for which experimental paradigms to employ, which helps address the corresponding issues I raised for the neuroscientific construct's validity vis-à-vis specifying cognitive capacities and experimental paradigms to study them.

Now that we have discussed how my theory of self-identification promises to help the neuroscience of gender, let us close by briefly considering why my theory is better suited to perform this work than other philosophical theories of gender on offer. As we have now seen from section five, my theory of self-identification (qua psychological gender) provides concrete definitional criteria which can be used to partition study groups consisting of concrete psychological components that can also specify the cognitive capacities involved in gender identity and so inform the design of experimental paradigms to study it. In contrast, the theory of gender identity (qua prescriptive gender ontology) as a 'self-identification' offered by Talia Mae Bettcher does not have the fine-grainedness necessary to stabilize the neuroscientific construct in the way my theory does. Although it introduces self-identification as involving beliefs and attitudes, it does not theorize the underlying psychology or offer criteria for delineating gender identities, instead advocating against theorizing these criteria and in favor of individual's first-person authority over their gender identities (Bettcher, 2009). The same problem concerning a lack of fine-grainedness applies to the theory of nonbinary gender identity (qua prescriptive gender ontology) offered by Robin Dembroff, which they call 'genderqueer', and classify as a 'critical gender kind'. Dembroff's theory also eschews theorizing the underlying psychology or criteria for delineating *specific* nonbinary gender identities, and instead analyzes them with the course-grained feature of destabilizing dominant gender ideology (specifically, what they call the 'binary axis'). This brings us to Jenkins's norm relevancy account of gender identity, which we already saw lacks fine-grainedness and so cannot perform the necessary work to stabilize the neuroscientific construct, since it could not explain how experiences of norm relevancy are translated into beliefs and desires about one's gender. These brief remarks about the other philosophical theories of gender show us that they are not as well suited to stabilizing the construct of gender identity as my theory of self-identification is. My theory of self-identification thus represents a unique philosophical contribution in addition to being a significant step forward in the multidisciplinary effort to stabilize gender identity.

REFERENCES

Bannerman, D., Good, M., Butcher, S., Ramsey, M., and Morris, R. (1995). Distinct components of spatial learning revealed by prior training and NMDA receptor blockade. *Nature, 378*, 182-186.

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

Berenbaum, S. A. (2018). Evidence needed to understand gender identity: Commentary on Turban & Ehrensaft. *The Journal of Child Psychology, 59*, 1244-1247.

Bettcher, T. M. (2009). Feminist Perspectives on Trans Issues. *Stanford Encyclopedia of Philosophy*.

Bettcher, T. M. (2014). Trapped in the wrong theory: *Rethinking* trans oppression and resistance. *Signs*, *39*, 383-406.

Caselles, E. L. (2021). Epistemic injustice in brain studies of (trans)gender identity. *Frontiers in Sociology*.

Diamond, L. M. (2020). Gender fluidity and nonbinary gender identities among children and adolescents. *Child Development Perspectives*, *14*, 110-115.

Dembroff, R. (2020). Beyond binary: genderqueer as gender critical kind. *Philosophers Imprint, 20*.

Eichenbaum, H., Stewart, C., and Morris, R. (1990). Hippocampal representation in place learning. *The Journal of Neuroscience, 10*, 1, 3531-3542.

Fausto-Sterling, A. (2021). A dynamic systems framework for gender/sex development: from sensory input in infancy to subjective certainty in toddlerhood. *Frontiers in Human Neuroscience*.

Feusner, J. D., Lidstrom, A., Moody, T. D., Dhejne, C., Bookheimer, S. Y., & Savic, I. (2017). Intrinsic network connectivity and own body perception. *Brain Imaging and Behavior*.

Geuss, R. (1981). The Idea of a Critical Theory: Habermas and the Frankfurt School. New York, NY: Cambridge University Press.

Gonzalves, T. (2020). Gender identity, the sexed body, and the medical making of transgender. *Gender and Society, 34*, 6, 1005-1033.

Gulgoz, S., Glazier, J. J., Enright, E. A., Alonso, D. J., Dunwood, L. J., Fast, A. A., Lowe, R., Ji, C., Heer, J., Martin, C. L., & Olson, K. R. (2019). Similarity in transgender and cisgender children's gender development. *Proceedings of the National Academy of Sciences*.

Irby, C. A., & Avishai, O. (2017). Bifurcated conversations in sociological studies of religion and gender. *Gender and Society*, *31*, 647-676.

Jacqueline, S. (2016). Construct stabilization and the unity of the mind-brain sciences. *Philosophy of Science*.

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

Jenkins, Katherine. (2018). Toward an account of gender identity. *Ergo*.

Jones, B. A., Bouman, P. A., Haycraft, E., & Arcelus, J. (2019). Gender congruence and body satisfaction in non-binary transgender people: A case control study. *International Journal of Transgenderism*, *2-3*, 263-274.

Manzouri, A., Savic, I. (2018). Possible neurobiological underpinnings of homosexuality and gender dysphoria. *Cerebral Cortex*.

Mikkola, M. (2017). Feminist Perspectives on Sex and Gender. *Stanford Encyclopedia of Philosophy*.

McElhinny, B. (2003), Three Approaches to the Study of Language and Gender. *American Anthropologist*, 105: 848-852.

McKitrick, J. (2007). Gender identity disorder. *Faculty Publications*.

Perry, J. (1998). Myself and "I". In Marcelo Stamm (ed.), *Philosophie in Synthetischer Absicht*, 83-103.

Perry, J. (2001). Knowledge, possibility, and consciousness. *Philosophy, 301*, 457-461.

Thagard, P. (2012). The self as a system of multilevel interacting mechanisms. *Philosophical Psychology*.

Thagard, P., & Wood, J. V. (2015). Eighty phenomena about the self: representation, evaluation, regulation, and change. *Frontiers in Psychology*, *6*.

Uribe, C., Junque, C., Gomez-gil, E., Abos, A., Mueller, S. C., & Guillamon, A. (2020). Brain network interactions in transgender individuals with gender incongruence.

Van Leeuwen, N. (2012). Perry on self-knowledge. In Albert Newman Raphael van Riel (ed.), *Identity, Language, and Mind: An Introduction to the Philosophy of John Perry*. CSLI Publications.

Wood, W., & Eagly, A. H. (2015). Two traditions of research on gender identity. *Sex Roles*, *73*, 461-473.

Zeman, D. (2020). Subject contextualism and the meaning of gender terms. *Journal of Social Ontology*, *6*, 69-83.

Zheng, R. (2018). What is My Role in Changing the System? A New Model of Responsibility for Structural Injustice. *Ethical Theory and Moral Practice*, *21*, 869-885.