

# UC Irvine

## UC Irvine Electronic Theses and Dissertations

### Title

Social Kind\*: Joint Intentional Mechanistic Kind

### Permalink

<https://escholarship.org/uc/item/34n4w4qk>

### Author

Cooper, Rachel

### Publication Date

2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,  
IRVINE

Social Kind\*: Joint Intentional Mechanistic Kind

DISSERTATION

submitted in partial satisfaction of the requirements  
for the degree of

DOCTOR OF PHILOSOPHY

in Philosophy

by

Rachel Katherine Cooper

Dissertation Committee:  
Professor Annalisa Coliva, Chair  
Associate Professor Katherine Ritchie  
Distinguished Professor Margaret Gilbert

2023



## **DEDICATION**

For

Bernadette

## TABLE OF CONTENTS

	Page
<b>ACKNOWLEDGEMENTS</b> .....	V
<b>VITA</b> .....	VI
<b>ABSTRACT OF THE DISSERTATION</b> .....	VII
<b>INTRODUCTION</b> .....	1
<b>CHAPTER ONE: AN EXPLANATORY CRITIQUE OF SOCIAL POSITIONAL ACCOUNTS OF THE METAPHYSICS OF SOCIAL CATEGORIES</b> .....	7
INTRODUCTION.....	7
THE DIALECTIC THAT LEADS TO SOCIAL POSITIONAL ACCOUNTS.....	8
<i>Unreflective Realism, Eliminativism, and Social Constructionism</i> .....	8
<i>Thin Constructionism and Dynamic Nominalism</i> .....	11
<i>Social Positional Accounts of the Metaphysics of Social Categories</i> .....	14
<i>Radical Social Constructionism, Conferralism, and Critical Realism</i> .....	17
PARTICULAR CONCERNS WITH SOCIAL POSITIONAL ACCOUNTS OF THE METAPHYSICS OF SOCIAL CATEGORIES.....	19
<i>Heavyweight Metaphysical View of Kinds</i> .....	19
<i>Explanatory Target</i> .....	29
<i>Ontological Collapse</i> .....	37
CONCLUSION.....	48
<b>CHAPTER TWO: SOCIAL KIND*: JOINT INTENTIONAL MECHANISTIC KIND</b> .....	49
DEBUNKING, CONSTITUTIVE SOCIAL CONSTRUCTION, AND THE APPEARANCE QUESTION.....	49
SOCIAL CONSTRUCTION AS MECHANISTIC EXPLANATION .....	53
<i>Metaphysical vs. Causal Explanation</i> .....	53
<i>Why Mechanistic Explanation?</i> .....	57
JOINT INTENTIONAL MECHANISTIC EXPLANATION .....	62
<i>Joint Intentional Mechanisms</i> .....	62
<i>The Content of the Shared Intention</i> .....	67
EXPLANATORY APPEAL AND ANSWERING THE APPEARANCE QUESTION.....	70
<i>Answer to the Appearance Question</i> .....	70
<i>Other Causal Accounts and the Importance of Joint Intentionality</i> .....	71
<i>A New Understanding of a Social Kind</i> .....	73
<i>Explaining the Distinction between Overt and Covert Social Construction</i> .....	75
WORRY CASES AND ANTICIPATING OBJECTIONS.....	77
<i>Tiger Objection</i> .....	77
<i>Too Cognitive Objection</i> .....	79
CONCLUSION.....	80
<b>CHAPTER THREE: THE VIABILITY OF CRITICAL ANTI-REALISM</b> .....	82
INTRODUCTION.....	82
THE VIABILITY OF CRITICAL ANTI-REALISM.....	83
THE CLAIM THAT CRITICAL REALISM IS <i>NECESSARY</i> FOR SOCIAL CHANGE .....	86
<i>Background</i> .....	86
<i>Strict Ontological Naturalism and the Error Theorist</i> .....	87
<i>The Focus on Constitutive Social Construction as a Response to Hacking</i> .....	89
<i>Contrast between Thin and Thick Constructionism</i> .....	90

<i>Critical Realism as a Response to the Radical Social Constructionist</i> .....	91
SKETCH OF CRITICAL ANTI-REALISM AND COMPARISON WITH CONFERRALISM .....	94
<i>Essentiality Conferred</i> .....	94
<i>Conferralism vs. The Joint Intentional Mechanistic Kind View</i> .....	97
APPLICATION TO A CASE: SEXUAL ORIENTATION.....	98
<i>The Main Accounts of Sexual Orientation</i> .....	99
<i>How the Conferralist Handles the Case</i> .....	101
<i>Various Objections to Social Positional Accounts in General and the Conferralist Account in Particular</i> ....	102
<i>The Result on the Joint Intentional Mechanistic Kind View</i> .....	107
HOW THE PROBLEMS OF CHAPTER ONE ARE AVOIDED.....	110
<i>Thoroughgoing Realism</i> .....	110
<i>Explanatory Target</i> .....	111
<i>Ontological Collapse</i> .....	112
RESPONSES TO WORRIES.....	113
<i>The Extensions of Social Group and Social Kind</i> .....	113
<i>The Role of Representation</i> .....	115
<i>Commitment to a Representation</i> .....	115
<i>Too Cognitive Revisited</i> .....	116
<i>Extra Ontological/Proliferation of Social Groups</i> .....	117
<i>Dynamic Nominalism and Radical Social Constructionism Revisited</i> .....	118
CONCLUSION .....	118
<b>CONCLUSION .....</b>	<b>120</b>
<b>BIBLIOGRAPHY.....</b>	<b>126</b>

## ACKNOWLEDGEMENTS

Thank you to Annalisa Coliva, Katherine Ritchie, and Margaret Gilbert. Thanks also to Marcello Fiocco, Lauren Ross, Sven Bernecker, and Allison Koslow. Thanks to Ásta. Thanks to the International Social Ontology Society, the Canadian Metaphysics Collaborative, the University of Groningen Summer School on Public Philosophy and Social Ontology, and the Philosophy of Social Science roundtable.

Thanks to my daughter Bernadette for her patience and understanding, to my parents and siblings for being there to support me, to my brother Matt for always being available to talk ideas, to my partner Clayton for his emotional support and encouragement, and to my friends for not forgetting about me while I mostly disappeared.

Thanks to my friends in the department, Rena Goldstein, Itzel Garcia, Nicholas Smith, Darby Vickers, Louis Doulas, and Matthew Dean. Special thanks to Matthew Dean and Deirdre O'Reilly for helping to look after Bernadette when needed.

Thanks also to the people at the Extended Day center and Children's center at UCI for giving me the peace of mind that Bernadette was in a nurturing environment after school.

**VITA**  
**Rachel Katherine Cooper**

2012	B.A. in Philosophy, San Francisco State University
2015	M.A. in Philosophy, San Francisco State University
2014-15	Graduate Teaching Associate, San Francisco State University
2018-21	Teaching Assistant, University of California, Irvine
2022-23	Teaching Assistant, University of California, Irvine
2022-23	Instructor of Record, University of California, Irvine
2023	Ph.D. in Philosophy, University of California, Irvine

FIELD OF STUDY

Social Metaphysics



## ABSTRACT OF THE DISSERTATION

Social Kind\*: Joint Intentional Mechanistic Kind

by

Rachel Katherine Cooper

Doctor of Philosophy in Philosophy

University of California, Irvine, 2023

Professor Annalisa Coliva, Chair

Some kinds that, upon reflection, seem to be in some sense “made up by us,” can nevertheless *appear* natural. For instance, there are reasons to think that races and genders lack a biological essence of the sort associated with paradigmatic natural kinds, and yet they are widely regarded as natural categories. This phenomenon of a kind appearing natural calls out for explanation. One kind of explanation that has been given is that such kinds are in fact unified by a social position. This dissertation critically examines this explanation and the metaphysical view implied by it. It also defends another way of explaining the phenomenon of interest, namely, by thinking about representations and certain *mechanisms* involving those representations.

In this dissertation I examined the question of how a kind that appears natural could be social. I studied the ways that the claim that some kind is *socially constructed* has been understood and evaluated prominent proposals for how to understand such claims while also considering existing critiques of those proposals. I studied how the notion of an *explanation* has typically been understood in various subfields, as well as the various ways one might understand the claim that some phenomenon is *distinctly* social. I came to the view that for a kind to be social, for a kind to be socially constructed in an important sense, is for the kind to be featured in

regularities that are explained in part by *joint intentional mechanisms* involving certain sorts of *covertly normative contents*. I argued that this proposal provides a more direct answer to the question I sought to answer, the appearance question, than does the prominent approach to understanding the notion of social construction and of social kinds while also avoiding some of what I take to be practical and theoretical shortcomings of that approach.

## INTRODUCTION

Social constructionists hold that there are *social kinds*: kinds that share a social property, i.e., a property that one possesses in virtue of something about other people. They disagree about what that something is (Ásta, 2017, p. 2).<sup>1</sup> The *debunking* social constructionist often suggests more specifically that the social kinds of interest are akin to social classes that one belongs to in virtue of being *taken* to have certain features, and where the properties one *actually has* are irrelevant to kind membership. Being taken to have some feature or features results in the conferral of a social status (Ásta, 2018: 44-45) or marks one for a subordinated social position in a (often hierarchical) social structure (Haslanger, 2012: 132-133).

A “debunking” project is one aimed at revealing that a kind that is widely regarded as natural is in fact social, or that a kind regarded as social is more deeply social than has been properly recognized (Haslanger, 2012: 132-136; Ásta, 2018: 36-37).<sup>2</sup> Debunking style social constructionists, or those who offer what are sometimes called *social positional theories* of the kinds of interest, often suggest that what makes such kinds distinctively social is that they are *constitutively socially constructed*. To say that a kind is *constitutively socially constructed* is to say that x cannot be defined without reference to social factors, or to say that x is made up in part by social processes, relations, interactions (Haslanger, 2012: 131; Ásta, 2018: 41).

The debunking social constructionist is concerned with uncovering the “cogs and belts” of oppression and it is common within the relevant literature to suggest that at least one

---

<sup>1</sup> Ásta writes, “All answers to the question what makes a kind a social kind appeal to something about subjects, but what that “something” is varies (Ásta, 2017, p. 2).

<sup>2</sup> For example, one might be surprised to hear an argument to the effect that gender is a social matter. One might be much less surprised to hear an argument to the effect that being a refugee or a widow is a social matter, but still surprised to learn that the term might refer to a “thicker” social position than one had previously recognized (Haslanger, 2012; Ásta, 2018).

important way of contributing to this task is to reveal that race, gender, and other social categories of interest to the debunker are *metaphysically dependent* on social structure.<sup>3</sup> Showing that such kinds are *dependent on us* in this way is meant also to show that the advantages and disadvantages that come with membership are contingent, alterable, and so demand justification (Ásta, 2018: 36; Haslanger, 2012; 136). If such categories were shown to be determined in a way that is completely independent of our thinking and practices, or to be tracking properties that are genuinely explanatory, i.e., outside of the *social significance* they possess, then a demand for justification of the relative advantage and disadvantage that comes with kind membership would be inappropriate (Ásta, 2018: 36).

The aim of this dissertation is to suggest reasons in favor of a reorientation of the debunking project away from such an approach—an approach of providing *real definitions* of particular social kinds—and toward answering the question: how can a kind that is plausibly partially or entirely socially constituted come to *appear* natural?<sup>4</sup> I make a case for thinking that this is the central question posed by the debunking social constructionist, and then I make a case for thinking that this question is not best answered with a *metaphysical* explanation but rather with a particular sort of *mechanistic* explanation. I provide an explanatory critique of what I term *social positional* accounts of the metaphysics of social categories.<sup>5</sup> It is an *explanatory* critique as opposed to an *ontological* one, insofar as the complaint made about social positional accounts has not to do with whether some category is rightly or wrongly counted as ontologically significant, but rather with whether the central question of the debunking social constructionist is

---

<sup>3</sup> The phrase “cogs and belts” comes from Ásta, 2018: 4.

<sup>4</sup> Kinds that, upon reflection, seem to be in some sense “made up by us” are sometimes the same kinds that are widely held to be biologically determined.

<sup>5</sup> I do not claim that I am *only* providing an explanatory critique, of the deflationary sort attributed to Carnap, where what is at issue is only “putting forward different proposals for how to talk” (Manley, 2009, p. 16). A case could certainly be made that there are significant metaphysical implications of the alternative proposal I end up offering.

best answered with a *metaphysical* explanation at all, let alone with the particular metaphysical explanation offered by the social positional theorist.

The phenomena of particular interest to the debunker, I argue, is when a kind appears natural but is in fact social. However, if we think of social constitution (a relation of metaphysical dependence) as the defining feature of social kinds, it seems we do not sufficiently single out the kinds of particular interest to the debunker. In focusing on countenancing real social positions, the social positional theorist does not properly highlight what is unique to cases of covert construction and so does not cover all of the relevant cases. I argue that there are kinds that have the special type of contingency, i.e., that are social kinds, that are not social positional, e.g., *people assigned female at birth, people who sincerely identify as women, bisexuals, autistic people*, and so on. I argue further that the social positional theorist includes cases that should not be covered by the notion of social construction. For instance, this characterization would lead us to include kinds like *money, president, judge, point guard*, etc. These seem plausibly constitutively socially constructed but seem to lack the same deep contingency as the covertly socially constructed kinds of interest. The social-positional account fails to single out the phenomenon of primary interest to the debunker—when a kind that is social appears natural—and so leaves certain epistemic benefits out of reach.

I argue that the social-positional theorist proposes an account of social construction that is both over and under inclusive, and that the mistake arises due to a failure to appreciate the role of *plural subject phenomena*.<sup>6</sup> In particular, I suggest that joint intentional phenomena of a particular sort figure in a *mechanistic* explanation of the regularities observed among members

---

<sup>6</sup> The term *plural subject* is introduced by Margaret Gilbert to refer to a given set of persons that are *jointly committed*. See for instance Gilbert, 2014: 6-9; Gilbert, 1989.

of paradigmatically social kinds. I argue that this relationship to a particular sort of joint intentional phenomena is what makes the kinds of interest distinctively social, rather than their being socially constituted. Although they may very well be partially socially constituted, I make a case for thinking that this is not what makes them social in the most important or especially interesting sense. The point to make is not that members of social kinds have nothing of importance in common beyond a social position, but that many of the regularities observed among them are explained in part by joint intentional mechanisms.

The social positional theorist is silent on—or at the very least underemphasizes—the relevance of such explanations, and I argue that this is to the detriment of the debunking project, the primary project the social positional theorist is aimed at advancing. That is because such explanations are required to reveal the radical type of contingency at play in the cases of primary interest to the debunker. In focusing on relations of metaphysical dependence, the social positional theorist might be seen as tethering the posited social positions to the “fundamental” in a way that might inadvertently make it harder rather than easier to take seriously the very mechanisms that debunkers aim to reveal.

The metaphysical explanation offered by the social positional theorist not only does not best answer the central question, in attempting to answer this question with a metaphysical explanation, the social positional theorist requires one to take on unnecessary metaphysical baggage—such as a heavyweight metaphysical view of kinds and in turn a thoroughgoing realism—needlessly alienating all but the metaphysical realist.<sup>7</sup> What if we could answer the central question in a way that could be adopted by realists and antirealists alike? The social

---

<sup>7</sup> Here I draw on arguments made by Elizabeth Barnes (2017) and Ron Mallon (2007).

positional theorist's insistence on capturing the explanatory power of social categories by positing something like a *social essence* forces one to take on unnecessary metaphysical baggage, such as a heavyweight metaphysical view of kinds. The social positional theorist's insistence on countenancing real social positions in turn seems to require a commitment to thoroughgoing realism, again unnecessarily alienating those reluctant to take on such metaphysical baggage. Another of my aims is to show that we can capture what the social positional theorist is after without what I term *bifurcation*, and further, that in avoiding such bifurcation we are better able to isolate the cases of primary interest.

The alternative I will offer is not necessarily at odds with the core insight of social positional accounts, and indeed I would like to count myself as engaged in the debunking project. It could be seen instead as a recommendation that we move away from the currently prominent focus on relations of metaphysical dependence, and toward a focus on isolating and describing the mechanisms responsible for the phenomenon I take to be of primary interest to the debunker, namely, when a kind that is social appears natural. I argue that taking such an approach allows us to accommodate those who recognize a special type of contingency at play in the cases of primary interest and yet who would not be moved toward a heavyweight metaphysical view of kinds or a thoroughgoing realism.

In the first chapter I give an overview of how what I term *social positional* accounts of the metaphysics of social categories come about, what they are a response to. I then offer three critiques of social positional accounts of the metaphysics of social categories. The first is that they require one to take on a heavyweight metaphysical view of kinds that is not necessary to capturing the central insight of the debunking social constructionist. The second is that they do not successfully single out the cases of primary interest to the debunking social constructionist,

namely covert cases of social construction. The third is that their focus on relations of metaphysical dependence makes them compatible with grounding frameworks, which I argue allow the phenomena of interest to be only *apparently* causal.

In the second chapter I argue that the central question of the debunking social constructionist is the appearance question and that a metaphysical explanation of the sort offered by the social positional theorist is not capable of answering the appearance question. I then provide a positive view for how to understand the notion of a social kind such that the appearance question is answered. In particular, I propose that a mechanistic explanation is required to answer the appearance question and that joint commitments with certain sorts of contents are the mechanism that explains regularities among the social kinds of interest.

In the third chapter I argue against the claim that critical realism is required for social change and make a case for the viability of critical anti-realism. I make clear the differences between the joint intentional mechanistic kind view and the conferralist view of the metaphysics of social categories, both of which I take to be forms of critical anti-realism that meet the aims of the debunking project. I then argue that the joint intentional mechanistic kind view has epistemic benefits that outweigh those of the conferralist view and respond to possible objections debunkers may have to my proposed view.

In the conclusion I briefly discuss another important upshot of the joint intentional mechanistic kind view, namely, that it suggests a method for changing the joint commitments that I have hypothesized partially explain how social kinds can appear natural.



# CHAPTER ONE: AN EXPLANATORY CRITIQUE OF SOCIAL POSITIONAL ACCOUNTS OF THE METAPHYSICS OF SOCIAL CATEGORIES

Abstract:

In this chapter, I give an overview of the dialectic that I take to lead to *social positional* accounts of the metaphysics of social categories. I then offer three critiques of social positional accounts of the metaphysics of social categories. The first is that they require one to take on a heavyweight metaphysical view of kinds that is not necessary to capturing the central insight of the debunking social constructionist. The second is that they do not successfully single out the cases of primary interest to the debunking social constructionist, namely covert cases of social construction. The third is that their focus on relations of metaphysical dependence makes them compatible with grounding frameworks, which I argue allow that the phenomena of interest are only *apparently* causal.

## INTRODUCTION

Debunking social constructionists want to reveal as social categories that are widely held to be natural, to reveal that certain *kinds* that are widely held to be natural are in fact social in some sense. A prominent approach to meeting this aim is to argue for a *social positional* view of the kind. Social positional accounts are those that recommend we understand social kinds that are paradigmatic for the debunking social constructionist, such as race and gender, as positions in a social structure, or as unified by a social property. In this chapter I raise some particular concerns with what I term *social positional* accounts of the metaphysics of social categories.

I will begin in section two by reviewing the debate that the social positional view arises from, focusing on unreflective realism, thin constructionism, dynamic nominalism, and radical social constructionism. In section three I provide three critiques of social positional accounts 1) social positional accounts needlessly require a heavyweight metaphysical view of kinds 2) social positional accounts fail to pick out the explanatory target of the debunking social constructionist and 3) social positional accounts are compatible with a grounding framework and so enable the view that the social phenomena of interest are only *apparently* causal.

## THE DIALECTIC THAT LEADS TO SOCIAL POSITIONAL ACCOUNTS

### *Unreflective Realism, Eliminativism, and Social Constructionism*

As has often been noted by those interested in the notion of social construction, the cases of primary interest to the social constructionist—namely, the cases of race, gender, and other human kind categories—have the following in common: that the default stance toward each is or has until relatively recent history been that of *unreflective realism* (Mills, 1998; Hacking, 1999; Haslanger, 2012; Mallon, 2004; Mallon, 2007, Ásta, 2018). The cases of primary interest are the subjects of what are sometimes referred to as *debunking* projects. The aim of a debunking project is to reveal as social a category that is widely held to be natural (Haslanger 2012; Ásta, 2018).<sup>8</sup> Indeed, one might think that *all* claims correctly classified as social construction claims are aimed at such debunking. Ian Hacking writes in 1999, “people begin to argue that X is socially constructed precisely when they find that: In the present state of affairs, X is taken for granted; X appears to be inevitable” (Hacking, 12). Two of the most notable debunking projects to date have been aimed at the categories of gender and race, both of which fit the foregoing description well. Charles Mills writes in 1998,

“for the past few hundred years... people have believed that there are natural biological differences among races and that these differences run deeper than mere phenotypical traits” (Mills, 46).

Sally Haslanger writes in 2000,

“It is always awkward when someone asks me informally what I’m working on and I answer that I’m trying to figure out what gender is. For outside a rather narrow segment

---

<sup>8</sup> This does not cover all cases of interest, however. Sometimes the category of interest is thought to be obviously social, so the aim is to reveal that it is social in a deeper sense than has previously been appreciated. Ásta writes “Not all debunking projects involve revealing a kind to be social that is widely believed to be natural. Sometimes a kind is widely believed to be social, so that is not the erroneous belief in question. Instead, the widely held but erroneous beliefs concern the nature of that social kind and the justification of the constraints and enablements that come with membership in the kind” (Ásta, 2018, p. 36). E.g., refugee, widow.

of the academic world, the term ‘gender’ has come to function as the polite way to talk about the sexes. And one thing people feel pretty confident about is their knowledge of the difference between males and females” (Haslanger 2012, p. 221).

These statements capture well the *unreflective realism* that is or has until relatively recent history been the default stance toward gender and race categories.

The *unreflective realist* about a human kind category likens it to a robust biological natural kind and so is likely to view the members of the category as sharing something like an *essence*.<sup>9</sup> An essence is typically understood to be an intrinsic and natural property or set of such properties that are individually necessary and jointly sufficient for membership in the category in question and which can explain the shared appearances, behaviors, and other observable properties among members of the category (Mallon, 2007, p. 148). To take an uncontroversial example, the essence of *water* is often held to be H<sub>2</sub>O: a molecule cannot be a water molecule if it is not composed of two hydrogen atoms and one oxygen atom.<sup>10</sup> When the essence is present, certain predictions are licensed, in this case, for example, that the substance will freeze at 32 degrees Fahrenheit. Social construction claims often start to arise when the discovery of such an essence for a putatively natural kind begins to look improbable. For example, take the cases of gender and race: Is there any one thing that all and only members of a particular gender or race share?

In the case of race, there simply is not a significant and explanatory commonality in genetic material that is shared by all and only members of the putative races as they are ordinarily divided (Appiah, 1996: 68; Mallon, 2007: 151; Haslanger, 2012: 235). It may be that there is a significant genetic commonality among relatively isolated groups over long periods,

---

<sup>9</sup> While we will see that realism about a kind need not entail essentialism about that kind, it is fair to say that realism about natural kinds is associated with essentialism about those kinds (Mills, 1998, p. 46).

<sup>10</sup> Another way it has famously been put is that in all possible worlds, water is H<sub>2</sub>O (Kripke, 1980).

such as the Amish, or Irish protestants, but again, these will not match with our ordinary ways of dividing people into races.<sup>11</sup> Furthermore, different characteristics are considered relevant for racial distinctions in different contexts: one may count as Black in Brazil but not in the U.S. (Haslanger, 2012: 235). Both realizations should lead one to deny the reality of race *as a biological kind*.

As can be seen in Ásta's (2018) discussion of the history of thought on sex and gender, while many post-Beauvoirean feminist theorists have focused on bringing into question the naturalness of *gender* rather than of sex itself—the very distinction of which can be attributed to Simone de Beauvoir's influential view presented in *Second Sex* (1949)—there is reason to question the naturalness of the division into two sexes as well.<sup>12</sup> Judith Butler argued that rather than understanding gender as “the social meaning of sex”, where sex is understood as a matter of biological fact, we should understand *sexual categories themselves* as “regulatory ideals” (Butler, 1993, p. 1).<sup>13</sup> Others do not go quite so far. Haslanger, for instance, admits of *objective types* that correspond to our sex categories, even while questioning the naturalness of gender. But in support of the stronger claim made by Butler, it is worth noting that recent work in biology shows that, for each biological trait we might point to as the defining feature of the sexes—for example, functioning genitalia, chromosomes, hormonal levels—we will be led to place people in different groups, depending on which method is used (Ásta, 2018: 71).<sup>14</sup> Here again we are forced to consider that sex may not simply be one of “nature's givens”.

---

<sup>11</sup> See Mallon, 2006, p. 533; Appiah, “Race, Culture, Identity” p. 72; Zack, *Philosophy of Science and Race*, p. 77

<sup>12</sup> Ásta, 2018, p. 55 notes that some scholars of de Beauvoir insist that this standard interpretation is erroneous.

<sup>13</sup> Ásta describes a regulatory ideal as “a prescriptive norm projected or posited by subjects, as opposed to, say, read off nature, so to speak” (Ásta, 2018, p. 61).

<sup>14</sup> See also Fausto-Sterling 2000a, 2000b; Roughgarden 2004; Callahan 2009.

While at one time the claim that some category is socially constructed may have been plausibly interpreted as the relatively modest claim that the category in question does not (as once believed) constitute a robust biological natural kind, it is now much more likely that the intended meaning of such a claim is stronger: that the kind is real, but not a biological kind (Mallon, 2004; 2006). In response to the *eliminativist* about the category, who reasons from the widespread acceptance (amongst experts, at least) that there is in fact no shared biological essence to be found among members of a category, to the conclusion that the kind does not exist at all, the social constructionist objects that it does not follow from the falsity of unreflective realism that the category or kind in question is not “real in other senses” (Mills, 1998, p. 47).<sup>15</sup> Indeed, one may find it perplexing that the eliminativist is able to jump sincerely from no-essence to non-existence, but I will not address this concern here. The point is that the social constructionist is one who moves instead toward the project of determining what we are *in fact* referring to in such cases, if not biological kinds (Mills 1998: 47; Haslanger 2012: 132-137).

### *Thin Constructionism and Dynamic Nominalism*

The social constructionist may adopt the view that the human kind in question is real in the sense of being, at the very least, *epistemically* objective (Mills, 1998: 58-59; Hacking, 1999: 22).<sup>16</sup> For instance, one can understand *thin constructionism* as referring to the view that a kind is social when it is *ontologically* subjective but *epistemically* objective; when the kind in question is plausibly considered ontologically significant (i.e., in some sense *real*) but this significance (i.e., reality) is rooted in something like what Charles Mills terms *intersubjective agreement* (Mills, 1998: 58). Just as I know that the \$1,200 in my bank account counts as

---

<sup>15</sup> For examples of eliminativist views of race see Appiah (1985) and Zack (1993).

<sup>16</sup> The term *epistemically objective* comes from Searle 1995, which is cited by Hacking 1999, but roughly the same notion is described in Mills 1998 as well.

payment for my rent due to intersubjective agreement about what counts as payment, the thin constructionist holds that we belong to the races, genders, or other putatively social categories that we do due to intersubjective agreement about what properties are relevant for membership. For example, in the case of race, the thin constructionist will say that *since* we have an intersubjective agreement that ancestry determines race membership, ancestry determines race membership (Mills, 1998). The thin constructionist about race or gender holds that there is an objective fact of the matter (of a sort) about what race or gender one belongs to, but that it is a sort of objectivity which is rooted in intersubjective agreement on Mills' view, or, as on Ron Mallon's view, in common knowledge representations of the category (Mallon, 2016: 67).<sup>17</sup>

On Hacking's *dynamic nominalism*, we again understand socially constructed kinds as those where the basis for a collection being a type or kind is non-objective, i.e., at best epistemically objective. A *thoroughgoing* nominalism refers to the view that "facts are consequences of ways in which we represent the world," that the world does not have an inherent structure that we discover, "that the world does not come with a unique prepackaged structure" (Hacking, 1999: 33 & 60). As Haslanger has put it, the nominalist may "grant that there is a distinction between types and random sets of things, but maintain that the basis for being a type is non-objective, that is, dependent on us" (Haslanger, 2012, p. 150). Sometimes nominalism is directed at one type in particular. When describing her *resemblance nominalism* about gender, Stoljar explains that on her view, women do not resemble each other because they are women, rather, what makes them women is that they resemble one another (Stoljar, 2011: 28; Rodriguez-

---

<sup>17</sup> Mills' thin constructionism is about race but can plausibly be applied to other human categories the thin constructionist deems social in the same sense.

Pereyra, 2008). For the nominalist, it is something about *us* that makes the collections we wish to refer to significant.

According to Hacking's dynamic nominalism about socially constructed kinds, a social kind is one that comes into being with the concept it. The concept of being such a person (a person of a particular kind) must become available, and to be a member one must intentionally act under a particular description (Ásta, 2018: 51). For Hacking, what makes a kind distinctively social is that it changes with the descriptions given of it. For instance, one might question whether there were homosexuals before the *concept* of being homosexual came into existence. We may retroactively recognize the behavior of certain individuals throughout history as homosexual behavior, according to current classifications, but there is room to question whether such individuals are properly regarded as *homosexuals*. That is because the kind, to the extent it is plausible to claim that there was one at all yet, *changed* with the introduction of the *concept* of the kind and the descriptions associated with it. There are surely many descriptions of homosexuality that would not fit the individuals throughout history who engaged in what we would now deem homosexual behaviors, who lived before we had the relevant concept. The dynamic nominalist holds that such individuals could not be *homosexual* without the concept being available to inform their self-conceptions and corresponding behaviors. For Hacking, a social kind is a description-dependent kind (Ásta, 2018: 52). His view is a *dynamic* nominalism about social kinds in that members of such kinds change in accordance with the available descriptions of them. Notably, this kind of "looping" is only possible with human kind categories.

Hacking focuses much of his discussion on the way that being categorized in a certain way shapes the behavior of category members: it does so by shaping one's self-conception and

thus one's behavior. Another focus of Hacking's discussion of social construction is non-inevitability. The contingent histories that led to certain idea constructions of, e.g., homosexuality, are non-inevitable: the idea constructions could have been different, and the corresponding self-conceptions could have differed accordingly, thereby shaping the behavior of those so categorized differently than they in fact did.

### *Social Positional Accounts of the Metaphysics of Social Categories*

The social-positional theorist goes a step further than the thin constructionist, and instead endorses a *thick* constructionism about socially significant categories. The *thin* notion of construction is the notion on which we admit the kind in question exists, but insist that the only sense in which the kind can be considered objectively determined is in the sense of being determined by what Mills calls *intersubjective agreement* (Mills, 1998: 58) or perhaps what Mallon terms *common knowledge representations* (Mallon, 2016: 67). The *thick* notion of construction favored by the social-positional theorist is the notion on which we again admit that the kind exists, but insist that there is *no independent fact of the matter* (intersubjective or otherwise) about which kind one belongs to, *outside the way one is viewed and treated*.<sup>18</sup> To belong to the kind is to be *taken* to have certain properties and be marked for certain kinds of treatment according to the background ideology (Haslanger, 2012: 132-133, Ásta, 2018: 2-3). The properties one *actually has* are deemed irrelevant, because the kind of interest to the social-positional theorist is the one composed of those constrained and enabled in particular ways due to perceived difference.<sup>19</sup> If one is regularly or for the most part taken to be white, then one is white. If one is regularly or for the most part taken to be a man, then one is a man. What makes a

---

<sup>18</sup> At the same time, however, there is a stark realism that comes with this view. The social positions and the social structures on which we find them have an existence quite independent of our thinking or knowing about them.

<sup>19</sup> Understanding the collection of interest in terms of shared constraints and enablements comes from Ásta. See Ásta's (2018) conferralist view for an important variation on the social-positional view.



kind social, thinks the social-positional theorist, is that it is unified by a shared social role: being taken to have some property or set of properties marks one for either a privileged or subordinated position in a (often hierarchical) social structure.<sup>20</sup>

On the thin constructionist view, there is a chance of being miscategorized. For instance, if there is intersubjective agreement on the single drop rule, then even if one is regularly taken to be white—i.e., even if one regularly *passes* as white—one can later be discovered to “in fact” or “really” be Black, when it is discovered that one has even a single Black ancestor.<sup>21</sup> According to the thin constructionist, it is possible to be something other than what you are taken to be. Not so for the thick constructionist. On the thick constructionist account, if this person who discovered this new information about their ancestry is taken on a regular basis to be white and is thus marked for a privileged position in the hierarchical social structure according to the background ideology, then they are white, and no amount of ancestral or other information will change this fact. To belong to a social kind is to occupy a position in a hierarchical social structure that is as real as rocks, trees, and semi-trucks speeding down the highway.

In response to Hacking’s dynamic nominalism, Sally Haslanger (2012: 125-137) responds that this notion of social construction does not capture what we need to capture with the notion of social construction. It is not simply that things could have gone differently, that, e.g., femininity and womanhood are understood and informed in certain ways because of historical developments that could have gone differently. What needs to be captured, according to Haslanger and other social-positional theorists, is that to be, e.g., a woman *just is* to be

---

<sup>20</sup> Some views that I will eventually include with those termed *social-positional* views do not in fact require that the relevant structure is hierarchical (e.g., Ritchie, 2020).

<sup>21</sup> For the constructivist, as opposed to the biological/unreflective realist “ancestry is crucial...because it is *taken* to be crucial, because there *is* an intersubjective agreement” (Mills, 1998, p. 58).

constrained and enabled in particular ways: it is about what one can and cannot do or get away with—an insight attributed to Simone de Beauvoir, when she wrote that “one is not born, but becomes a woman”—and the social positional theorist argues that one need not be *aware* of anything in particular in order to be so constrained or enabled (de Beauvoir, 1949; Haslanger, 2012; Ásta, 2018: 42-45).

Such constraints and enablements are dictated by ideology: what Haslanger terms *cultural techne*, and what Ásta terms a *social map* (Haslanger, 2017; Ásta, 2018: 115-126). According to the social-positional theorist, gender, race, and other social kinds are not social kinds in virtue of merely being socially caused or influenced, but rather in virtue of being *constitutively* socially constructed. To belong to the kind is just to occupy a position in a hierarchical social structure (Haslanger, 2012), or to have constraints and enablements conferred onto one upon being taken to have some relevant grounding property (Ásta, 2018). Social kinds are thus held to be unified by a social position or a social property. The social-positional theorist prioritizes being able to refer to the realities of being placed in a particular category, realities that in no way dependent on our awareness. There is a stark realism that is both encouraged by the social-positional theorist and which falls out of the standard articulation of the view.<sup>22</sup> It counterintuitively becomes a very mind-independent matter, who belongs to which kind, in that the kind is made up of just whoever is objectively occupying the real social position in a real social structure. Even if a white man discovers his black ancestry, intersubjective agreement about the importance of one’s ancestry does not change the facts about what privileges he enjoys due to being *taken* to be white.

---

<sup>22</sup> An exception to this might be Ásta’s (2018) conferralist view, which involves changing contexts and which seems to be compatible with a general conferralism about essences.

### *Radical Social Constructionism, Conferralism, and Critical Realism*

The *radical* social constructionist holds that all distinctions are made according to our interests, and so do not capture “nature’s joints”. The radical social constructionist is likely to reject the idea that “gender is the social meaning of sex,” because *sex itself* is constructed according to our interests. It is not that sex is one of nature’s givens, and gender is the social meaning of having a certain biologically determined sex. We fashion the sex binary to meet certain ends of ours. If sex is already constructed, it is of little use to even distinguish between sex and gender (Butler, 1990: 7; Hacking, 1999: 8-9). One can also find the claim, which Haslanger attributes to Catherine MacKinnon, that “the power that has determined gender categories is the *same power* that has determined all categories” (Haslanger, 2012, p. 78) and a corresponding skepticism about sorting things according to their supposed natures or essences, perhaps even when those essences are qualified as *social* essences, which one might understandably view the social positional theorist as proposing we do. Such a radical social constructionism may sound extreme, however, we should note that it is not just the radical social constructionist that is suspicious of our ability to sort things according to their inherent natures. Philosophers of science have doubts at the very least about the need for necessary and sufficient conditions for kind membership (Mallon, 2007).<sup>23</sup> Indeed the more standard position on kinds from a scientific point of view may be a type of *conferralism* about them.

Conferralism is a position defended by Ásta (2018) that is in the same spirit as the social positional view and yet which avoids the apparent metaphysical realism of that view. Ásta develops a view that is compatible with the idea that all essences are conferred (Ásta, 2008; Ásta,

---

<sup>23</sup> One may argue, as Rebecca Mason (2021) has, that necessary and sufficient conditions are not an essential part of the essence concept.

2013). It stays in agreement with the social-positional view in that social kinds are indeed unified by social roles we occupy as a result of being *taken* to have certain features. But it is unique in allowing that all essences could be conferred, and in accommodating the radical social constructionist in certain ways.<sup>24</sup> It can avoid the Butlerian objection, for instance, by understanding sex itself as a social or conferred property. Conferralism is a framework for understanding social kinds as kinds where being taken to have a grounding property (e.g., certain genitalia) results in the conferral of a social property (e.g., sex), and the corresponding constraints and enablements. However, if one were to take on conferralism about essences in general, or even take on the context dependence of the conferralist view, one may worry that the distinction between social and natural kinds again becomes unclear and furthermore, that the account is less well equipped than standard social-positional accounts (namely metaphysically realist ones) to account for the *systematicity* of the phenomena of interest.<sup>25</sup>

The social-positional theorist claims something more radical than both the thin constructionist and the dynamic nominalist but does not want to go as far as the radical social constructionist, or to be associated with antirealist or idealist views. The proponent of the standard social-positional view will tend to insist that kinds unified by a social property are nevertheless real kinds, that social kinds are no less *real* for being social. The social positional theorist is likely to stress that they do not endorse moving from the idea that *some* kinds are determined socially to the idea that *all* kinds are so determined and to insist on the importance of holding that there are kinds that are not social. There are kinds that are determined completely mind-independently, and retaining this contrast, the social positional theorist holds, is essential to

---

<sup>24</sup> As will be discussed in chapter three, conferralism about essences is the view that thought experiments at best tell us about our conceptual commitments (Ásta, 2008).

<sup>25</sup> This is in part due to the role of context in Ásta's account. This worry is addressed in Ásta, 2018: 127-128.

social change. There is a spectrum, within the real, from natural to social (Haslanger, 2012). We are capturing nature's joints when referring to social kinds, but they are joints we *made* (Barnes, 2017).

## PARTICULAR CONCERNS WITH SOCIAL POSITIONAL ACCOUNTS OF THE METAPHYSICS OF SOCIAL CATEGORIES

### *Heavyweight Metaphysical View of Kinds*

Prominent accounts of the metaphysics of social categories recommend that to belong to a socially constructed kind—to be a member of what is commonly referred to as a *social kind*—is just to occupy a position in a (often hierarchical) social structure, or to otherwise instantiate a social *property* or status (Haslanger, 2012; Ásta, 2018; Barnes, 2017, Ritchie, 2020). This is a move that allows us to acknowledge the realities of belonging to a socially significant category, without unfoundedly countenancing a mind-independent *essence* had by all and only those belonging to the category (Mills, 1998; Mallon, 2004; Haslanger, 2012). Occupying the position that one does is often said to be a matter of the properties one is *taken* to have, rather than a matter of the properties one *has*.<sup>26</sup> Many philosophers who we might classify as *debunkers* take this point to be of the utmost importance, as it allows us to reveal the radical contingency of occupying the position one does, and thus also of the privileges and disadvantages one accordingly receives (Haslanger, 2012; Ásta, 2018). Thus, the relevant kinds are sometimes argued to be unified by a *social* property rather than a natural one, or may even be said to share a relational, mind-dependent essence.<sup>27</sup> The essence or property in question is not however thought

---

<sup>26</sup> At times I will refer to this as *perceptual individuation* for convenience.

<sup>27</sup> For instance, see Mason (2016) for an explicit reference to relational essences. See also Mallon (2016) on the possibility of relational-property-cluster kinds.

to be any less *real* as a result of being so dependent, nor is the kind that is so unified. It is simply *not the thing we thought it was* (Haslanger, 2012).

Much of the work in the humanities and the social sciences dedicated to understanding human categories, and the ascriptions of the corresponding kind concepts to persons, shares a common commitment to doing so in a *non-essentialist* way (Mallon, 2007, p. 146). To be committed to non-essentialism about a category is to be committed to the view that instances of the category in question share no common essence. An *essence* is (traditionally) supposed to be a property or set of properties that are simple, that are individually necessary and jointly sufficient for membership in the category, that are intrinsic (nonrelational) features of the instances of the category, and that explain observable properties characteristic of the category (Mallon, 2007, p. 148). As Mallon points out, we can divide non-essentialist approaches to human kinds into two types: those that deny there are simple necessary and sufficient conditions for membership in human kind categories (call these the skeptical anti-essentialists) and those that deny that there are intrinsic of natural features shared by all and only the members of human kind categories (call these constructionist anti-essentialists).

Essentialist views have a notoriously bad name in social theory.<sup>28</sup> But Haslanger and other social-positional theorists, if they can truly be said to posit essences at all, only posit a *relational* essence: the claim is simply that the kind cannot be defined without reference to social factors. An essence has historically been understood as an intrinsic property or set of properties that are individually necessary and jointly sufficient for membership in a category, that may be unobservable, but which are typically thought to explain observable properties characteristic of

---

<sup>28</sup> “The invocation of such non-essentialism is ubiquitous in social theory, and non-essentialism is widely considered an important constraint that must be met” (Mallon, 2007, p. 146)

members of the category (Mallon, 2007, p. 148). One might think an *extrinsic* essence is no essence at all. However, there is something to this characterization, as it is claimed that the only thing shared by all and only members of the relevant category is a social position, and that if it came about that no one occupied the relevant social position, the kind would cease to exist.

It is not the case that every debunking social constructionist will endorse the view that the *essence* of whatever putatively social kind of interest is a social position, or that there need be an essence shared by members of the kind at all, or that the *only* thing all and only members of a social kind share is a perceptually individuated social position.<sup>29</sup> Some may allow that the individuation conditions for membership are multifaceted, or even that there is a fact of the matter, outside of how one is viewed and treated, about to which categories one belongs.<sup>30</sup> They will, however, tend to insist that the only other sense in which membership can be seen as objectively determined is in the sense of being determined in part by something like what Mills refers to as *intersubjective agreement* (Mills, 1998: 58). This could be seen as a requirement on a position still being considered *constructionist*. It is nevertheless fair to say that the related—albeit more extreme—position, that social kinds are unified by a perceptually individuated social position, has become increasingly prominent in recent years. And one way of interpreting *that* claim is as the claim that the *essence* of a paradigmatic social kind is a social position.

Appealing to the intuitions of the *semantic externalist*, the social positional theorist suggests that the kinds to which the term *social kind* is most appropriately applied are those kinds that are best understood as perceptually individuated social classes. The debunker suggests

---

<sup>29</sup> When I say that some debunking social constructionists hold that social kinds are *perceptually individuated*, I mean that they hold that belonging to the kind is a matter of being *taken* to have certain traits rather than actually having them. Of course, one may object that perceiving and taking to be are importantly different, however, having this discussion here would take us too far astray.

<sup>30</sup> See, for example, Mills (1998), Mallon (2007), Ritchie (2020).

that the social kinds of interest are akin to social classes that one belongs to in virtue of being *taken* to have certain properties, and where the properties one *actually has* are irrelevant to kind membership. That, e.g., race and gender are best understood as social positions that one occupies in virtue of being taken to have certain traits, and this is what it is for a kind to be social. Social kinds—understood in the prominent way, as perceptually individuated social classes—might be interpreted as sharing a *social essence*: an essence composed of a social property, that all and only members of that kind share. Debunkers may typically distance themselves from talk of social essence, however, at least on the face of it, it would appear to be a possible consequence that falls out of the view. Which is to say, I wonder if appealing to *semantic externalism* is really conducive to this distancing, given its close connection to *scientific essentialism* and the way it is combined with an insistence on metaphysical realism, or *critical realism*.<sup>31</sup>

The debunking social constructionist often makes a point of questioning the idea that objectivity, where this is taken to mean mind-independence, is the criterion for being real. They will say that it is not even helpful to cast the problem as a distinction between natural (real) and social (unreal) kinds (Haslanger, 2012: 213-215). It is more accurately cast as a spectrum from non-social to social, *within the natural—within the real*. For Haslanger, the point of the claim that some kind is socially constructed is not to say that it is *not real*, or *not natural*, just that it is *not the thing we thought it was*. To make sense of how this could be true, she points to the very influential theory of meaning termed *semantic externalism*: just as we could be deeply wrong about the meanings of natural kinds terms such as *water*, *beech*, *elm*, or *arthritis*, and still successfully refer, we are in fact deeply wrong about the meanings of certain social kind terms

---

<sup>31</sup> Haslanger discusses the views of scientific essentialism put forth by Putnam (1975) and Kripke (1980), while distancing herself from the “naturalistic and essentialist assumptions” involved (Haslanger, 2012, p. 134)



such as race and gender terms, even though we successfully refer to races and genders on a regular basis (Haslanger, 2012, p. 134). They are not kinds that are unified by starkly physical properties (e.g., genetic difference of anatomical difference, say) after all, but rather by our social practices. On Haslanger's account, this in no way implies that the social kinds are not *real*, or not *natural*, because they still supervene on, or are constituted by, some natural property or properties—just not on the physical characteristics of the person as we have thought, but rather on social structure.

As Haslanger notes, Putnam and others who put forth views on scientific essentialism, semantic externalism, and related views, in discussing the use of natural kind terms, assumed that the relevant experts who determine what the paradigms share would be natural scientists. They held that the kind sought by these experts would be the *essence* of the paradigms (Haslanger, 2012, p. 134). Haslanger suggests that we might, in the cases of interest to the debunker, instead defer to “certain social scientists”. If we are understanding the view as a variation on scientific essentialism, we might wonder who exactly the right social scientists are. If we are understanding the view as a variation on Tyler Burge's (1979) *social externalism*, on which the meaning of our terms are constitutively determined in part by social factors, we may wonder about the plausibility of Haslanger's suggestion that social theorists are the relevant experts. In either case, it is not clear the experts we choose will successfully fix the reference. If we could find out who successfully fixes the reference, would they have the result that is proposed by the social positional theorist? To this, Haslanger may respond that in the case that they do not, we should reject the terms and introduce new ones that pick out the collections of interest. However, it is worth thinking carefully about whether this works.

It is important to note that the social-positional theorist, in proposing something like a social essence, is not proposing that it is part of the essence of the *individual members* of the kinds in question to be subordinated or privileged. The social-positional theorist argues that “after the revolution” there will be no, e.g., Black people or women, as there will be no one subordinated or privileged on the basis of being taken to share ancestral links to a geographical region or a role in sexual reproduction respectively, but there will of course be individuals to which the relevant terms would have previously been aptly applied (Haslanger, 2012, Ch. 7; Ch. 9). It is therefore best understood as a kind-level essence (Mallon, 2007). Still, we may question the plausibility of social externalism in these cases.

However, as Ron Mallon (2007) has argued, non-essentialism is far too weak a constraint to do the work that social theorists want it to do. Namely, it is not able to block pernicious generalizations about members of the categories in question. He argues that the focus on non-essentialism results from the mistaken assumption that generalizations about a category require *simple necessary and sufficient conditions* for category membership. He does this by considering what philosophers of science take to be sufficient for a category to be considered as constituting a kind and thus for licensing generalizations and predictions about things that fall into that category. They do not require simple necessary and sufficient conditions. They also do not require that the properties shared be intrinsic properties. Indeed, non-essentialism is also a widespread position held by philosophers of science toward a variety of categories held to nevertheless be rightfully considered natural kinds. *Species, inflation, schizophrenia, money*, and so on are concepts that are considered useful for explanation and induction (Mallon, 2007). Indeed, they are seen by the relevant sciences as corresponding to natural kinds. Yet, it may very

well be the case that they lack necessary and sufficient conditions for membership, that they are unified by relational properties, or that their existence is socially caused.

I would like to point out that these very considerations also show that going the other route and actually positing *social* essences cannot do the work of blocking pernicious generalizations either. Surely it allows for generalizations about those that occupy the posited social position, and that is part of the point. Social constructionists are concerned to reveal the *real source* of the inductive potency of the kind, rather than to deny the reality of the kind. But I do not think it has been properly appreciated that this move nevertheless does not rule out generalizations about other clusters of properties with significant overlap in extension with the posited social position. That is because positing social positions does not rule out generalizations about those with a cluster of anatomical features, or a sincere self-identification with a particular gender, and so on. Here I have only aimed to point out that on one conception of essence, social position theories are committed to such things as essences. And to point out a weakness of the theory given that essences are a heavyweight metaphysical commitment that seem not to be required, i.e., they do not do the work they are meant to do.

The social positional theorist suggests that we are required to commit to the existence of something like a social essence in order to make true our generalizations about, e.g., race and gender kinds. However, not even paradigmatic natural kinds—e.g., physical or biological kinds—are thought by philosophers of science to share an *essence* (e.g., species). Rather, they are thought to share a cluster of properties about which we can make reliable predictions (Mallon, 2007). Thus, showing that there is no biological essence shared by members of the kind

does not block generalizations about individuals with the relevant grounding or target property.<sup>32</sup> It seems it has not been properly appreciated that these very considerations show that going the other route and actually positing *social* essences cannot do the work of blocking what Ron Mallon (2007) has referred to as *pernicious generalizations* either. It allows for generalizations about those that occupy the posited social position, and that is the point of positing such positions: social constructionists are concerned to reveal the *real source* of the inductive potency of the kind, rather than to deny the reality of the kind. This move nevertheless does not rule out pernicious generalizations about property cluster kinds that share significant overlap in extension with the posited social positional kind. For example, positing social positions does not rule out generalizations about those with a cluster of anatomical features, or a sincere self-identification with a particular gender.<sup>33</sup>

Furthermore, if we are realists about perceptually individuated social positions, and the social structures on which they fall, then it seems we are forced into realism “all the way down”. It would be odd to be realists about social structure and social positions, but anti-realists about tables and chairs, say. Elizabeth Barnes (2017) makes a similar point when discussing the emphasis on *fundamentality* in much of contemporary metaphysics, and the tension between such an emphasis and views like Haslanger’s *critical realism*. Barnes writes “...a fundamental theory that quantifies only over, say, atoms in the void and social structures just looks like a bad theory—the existence of social structures looks like magic” (Barnes, 2017, p. 18). We might see this worry as the inspiration for Jonathan Schaffer’s (2017) suggestion that we understand social construction claims as *grounding claims*. He recommends that to be socially constructed is to be

---

<sup>32</sup> The language of *grounding properties* comes from Ásta’s (2018; 2013; 2011) *conferralist* framework on which a kind is social if it is unified by a conferred social status (e.g., gender) that consists in constraints and enablements, and that results from being taken to have some grounding property (e.g., sex assignment).

<sup>33</sup> See Jenkins (2016), on the inclusion problem and the importance of a gender concept based on identity.

grounded in social patterns (which somewhere down the line are grounded in atoms and void). Later, I will raise concerns about the fact that social positional views do indeed seem to be easily made compatible with the grounding framework. For now, I would simply like to point out that positing relational essences implies a thoroughgoing realism that is not required to make sense of how the kinds of interest to debunkers are explanatory, and so needlessly alienates all but the critical realist.

Such an account requires a heavyweight metaphysical view of kinds (and in turn a thoroughgoing realism), but this move is not actually *necessary* to making the point that the social-positional theorist ultimately wants to make. Social essences are not required to make true generalizations about race and gender kinds. Not even natural kinds are thought to require essences in order to be genuinely explanatory. If the point of positing a social essence is to show how a social category can be explanatory, then it can be shown to be a superfluous posit, as essences are not required to make a category explanatory. The social-positional theorist often suggests that we are required to commit to the existence of something like a social essence in order to make true generalizations about, e.g., race and gender kinds. However, not even natural kinds are thought by philosophers of science to share an *essence* (e.g., species). Rather, they are thought to share a cluster of properties about which we can make reliable predictions (Mallon, 2007). It thus needlessly alienates all but the metaphysical realist, and so blocks widespread adoption of the core insight of constructionism. Remaining ontologically neutral would allow for more widespread adoption of the core insight of social constructionism.

Here I have only aimed to point out that on one conception of essence, social positional theories are committed to such things as essences. And to point out a weakness of such a theory given that this commitment does not do the work it is supposed to do and thus needlessly

alienates all but the metaphysical realist. Relational essences imply a thoroughgoing realism that is not required to make sense of how the kinds of interest to debunkers are in some sense made up by us and nevertheless explanatory, i.e., kinds. I have argued that the social positional account needlessly alienates all but the metaphysical realist. I made a case for thinking that such an account requires a heavyweight metaphysical view of kinds—and in turn a thoroughgoing realism—that is not actually necessary to making the point that the social positional theorist ultimately wants to make: that social practices explain regularities among members of socially significant categories.

It might be that one could make the notion of essence very innocuous. Rebecca Mason (2016) has made a case for thinking that essences need not be understood as a set of necessary and sufficient conditions. She holds that the essential properties of a kind are simply the properties that individuate a kind, and so the essential properties of a kind can be “clusters of properties that are contingently but reliably coinstantiated because they are held in homeostasis by one or more causal mechanisms,” also known as Homeostatic property clusters (Boyd, 1999; Mason, 2016). One might think that if essences are made sufficiently innocuous, essentialists and non-essentialists may no longer disagree on anything of importance. But even then, proponents of anti-essentialism—a controversial view but a view that people have—would nevertheless be unable to accept the social-positional theorist’s proposal. The proposed theory would still needlessly alienate those who do not wish to take on essences, or the thoroughgoing realism that is implied by a commitment to real social positions. The account needlessly alienates anti-essentialists and those who favor thin constructionism about kinds in general.

What if we could answer the central question in a way that could be adopted by realists and antirealists alike? I will argue that a thin constructionist about kinds in general could still

make sense of the crucial difference between social categories and natural ones. The difference namely between the type of contingency at play in the regularities in which said categories are featured. Theoretically, it seems desirable that both social and natural kinds could be understood as individuated in the same way, perhaps as Boydian homeostatic cluster kinds (Boyd, 1999), that feature in useful generalizations. Practically speaking, the crucial insight is more widely available if it can accommodate those opposed to metaphysical realism and/or essentialism of any sort. The goal, then, should be to uncover the mechanisms that make the relevant clusters explanatory and reliably predictive despite their being socially rather than naturally significant clusters. The way to capture the radical contingency of the regularities observed among members of social kinds is to focus on uncovering the mechanisms that are responsible for homeostasis in the cluster kinds suspected of being explanatory as a result of something other than the shared cluster of properties itself. It is unclear what value there is in calling this cluster an *essence*.

### *Explanatory Target*

It is not just that if we can accommodate those other than the metaphysical realist, we should. I hold the view that there are many kinds that share a cluster of perfectly ordinary properties—as opposed to a social property or relational essence—that are nevertheless social kinds. The social positional account is both over- and under- inclusive when it comes to picking out the kinds of primary interest to the debunker. Focusing on relations of metaphysical dependence such as social constitution does not enable one to single out all and only the cases of interest to the debunker. The kinds of most interest to the debunker and the kinds that are socially constituted are not plausibly coextensive. This suggests that the social positional account leaves certain epistemic benefits out of reach, such as an awareness of what it is that is in fact shared by

the kinds of primary interest to the debunker, i.e., kinds that seem to feature in regularities that are contingent in an especially interesting and important sense.

Some cases that intuitively feature in the same radically contingent type of regularity of interest to the debunker that are not plausibly entirely socially constituted/unified by a social position are *those assigned female at birth, individuals who sincerely self-identify as women, autistic people, bisexuals*. Focusing on the first two, plausibly many of the regularities observed among those *assigned female at birth* and among those *who sincerely self-identify as women* are non-inevitable in the relevant important sense. They are explanatory categories insofar as we can make reliable predictions about those who share the relevant cluster of properties, but their explanatoriness does not stem from the cluster of ordinary properties shared by members alone. Thus, they are intuitively social kinds but not social positional. As we have seen, such categories do not require essences, let alone social essences, to be explanatory categories. It seems they ought also to be counted as social kinds in the sense relevant to debunkers, whatever terms we use to pick them out. And yet, it does not seem that such collections are social kinds in virtue of picking out perceptually individuated social positions. Those *assigned female at birth* self-identify in a wide variety of ways and this impacts the ease in which they can navigate the practical situations in which they find themselves in equally various ways. Those *sincerely self-identified as women* have lived experiences that vary widely and come up against varying degrees of resistance to their self-identification. However, in both cases, there are indeed shared features, and there are indeed genuinely explanatory categories to refer to. And in both cases, these genuinely shared features do not come close to providing the whole story of what makes such categories inductively potent.



Or take the case of *bisexual women*. There is a higher rate of domestic violence against bisexual women. This does not seem to be explained by reference to intrinsic features of bisexual women, but neither does it seem plausible to say that to be a bisexual woman is to occupy a social role, or that the only thing all and only bisexual women share is a perceptually individuated social position. That a woman is bisexual is largely an invisible trait. It is thus highly dubious both that all the individuals in this category are significantly similarly situated socially *and* that there is *nothing else* that all and only those belonging to the category have in common.

Some cases that are plausibly entirely socially constituted/unified by a social position but do not intuitively feature in the same radically contingent type of regularity of interest to the debunker are *united states president, judge, point guard, class president, librarian*, etc. Plausibly they are categories that are unified by a social role or status, but they are not plausibly kinds that feature in regularities with the relevant sort of contingency. There are not systematic regularities of the sort of interest to the debunker that need to be explained among linebackers or vice presidents.<sup>34</sup> Even if one wanted to make the case that there were such systematic regularities, what would need explaining is not the appearance that such regularities are naturally occurring and inevitable. I have suggested that the phenomenon of particular interest to the debunker is when a kind that appears natural is in fact social. However, there are plenty of plausibly socially constituted kinds that are not social in this covert sense that debunkers are interested in. Thus,

---

<sup>34</sup> As will be discussed in chapter two, I am happy to admit that these are social kinds insofar as they feature in regularities that are explained by joint intentional mechanisms. It might be the sort of content of the joint intention that differs in covert and non-covert cases. The point is that they are not interestingly social simply in virtue of being constitutively socially constructed.

this characterization appears to be too broad to capture what debunkers want to capture with the notion of social construction.

The social positional theorist proposes that social kinds are constitutively socially constructed kinds, kinds that cannot be defined without reference to social factors. However, it is conspicuous that this proposal does not single out covert cases of construction, i.e., the cases of particular interest to the debunking social constructionist. We might not want to take on essences in general or social essences in particular for theoretical reasons, or for ethical reasons, but even if they are made perfectly innocuous essences (not understood as necessary and sufficient sets of conditions, not understood as intrinsic, and so on), kinds with social essences do not capture all and only the kinds of interest to the debunker. The same special type of contingency can be at play whether the real definition of the kind is a social position or not. Further, there are kinds or categories that plausibly best understood as unified by a social role that do not involve the special type of contingency of interest to the debunker. Because the social positional account does not properly highlight the specific mechanisms at play in covert construction, it includes irrelevant cases and fails to include relevant ones. Thus, it leaves certain epistemic benefits out of reach, such as awareness of what it is that the cases of primary interest to the debunker in fact have in common.

Not only are social essences not necessary to answering the central question of the debunker, the proposed analysis also does not cover all and only the cases of primary interest to the debunker. The account is both over- and under- inclusive, and this shows that it does not properly highlight what is unique to the cases of primary interest, namely covert cases of construction. An account of socially constructed kinds should cover all the relevant cases. Thus,

even if social essences are made perfectly innocuous, they do not serve as an analysis of the notion of a social kind the debunker is interested in.

For example, there is plausibly more than one concept *woman*. Plausibly, the regularities observed among those *assigned female at birth* (henceforth AFAB) and those *who sincerely self-identify as women* are non-inevitable in the relevant important sense. Thus, they are social kinds but not social positional. Such categories do not require essences, let alone social essences, to be explanatory categories.

To make the case that the social positional account does not cover all relevant cases, consider that one might allow that there is more than one concept *woman*. For example, Katherine Jenkins (2016) has argued that if we are to accept Haslanger's proposal that there is an important concept of *woman* that picks out those who occupy a subordinate social position, we will still need, in addition to this, a concept of *woman* that picks out those who sincerely self-identify as women. Furthermore, it seems that we could make many true generalizations about those who sincerely self-identify as women, and that many of the regularities observed among those who so identify would be non-inevitable in the relevant important sense.<sup>35</sup> This seems likely to be true of those *assigned female at birth* as well. I do not want to argue that we need a concept of *woman* that picks out those assigned female at birth. Rather, my point is that it seems a case could be made that the two collections of individuals mentioned, those who sincerely self-identify as women, and those assigned female at birth, ought also to be counted as social kinds,

---

<sup>35</sup> The relevant important sense is the systematic predictability of the property cluster that is not explained by the having of the cluster itself. It is not simply that the regularities among members are the result of contingent historical developments that could have been different.

whatever term we use to pick them out. And yet, it does not seem that such collections are social kinds in virtue of picking out a perceptually individuated social position.

The point is that, if we pick out a collection of individuals with a significant overlap in extension—to the Haslangerian social kind, *woman*, for example—and make generalizations about them, they will be true generalizations. But in each case, the generalizations will not require essence, let alone social essences. And in every case, it seems we need to point out that many of the regularities observed are non-inevitable in the relevant important sense. Positing social essences does not block pernicious generalizations about people who sincerely self-identify as women or people assigned female at birth, for example. That is because not only are the generalizations permitted, they also do not have the same extensions as the social-position kind.

Intuitively there are socially constructed kinds that are not unified by anything like the social essence that is posited in the cases taken to be paradigmatic. For instance, in the case of *woman* there is posited a subordinate social position that unifies members of the kind and explains how such a heterogeneous collection could constitute an explanatory category. But in the cases just mentioned it is not clear that there is a social position shared by all individuals in the collection and yet it does seem clear that many of the generalizations that are true of the collection are non-inevitable in a sense that should be captured by the notion of social construction. In other words, the proposed characterization of social construction as social constitution does not seem to cover all the cases of primary interest to the debunker. Social essences can explain some things. But people assigned female at birth and people who sincerely self-identify as a particular gender are social kinds but not social positional. If the aim is to

answer how a kind that appears natural could be social, then we should cover all such cases. But social positional essences do not cover all the cases of appearing natural but being social.

Then there are examples that are less frequently discussed in the literature but seem clearly to be featured in regularities that involve the same sort of radical contingency. What about bisexuals? Autistic people? These categories are even less plausibly unified by a social position but seem clearly to feature in the importantly non-inevitable sort of regularities. Take the example of bisexual women. There are robust regularities among members of such kinds that cannot be explained simply by reference to their shared features. But belonging to such kinds does not seem to be a matter of what features one is taken to have. The fact that one belongs to one of these kinds might be largely invisible. Mills (1998: Ch. 3) makes an important point when discussing those who can “pass” as something other than what they “really” are, about the difference one’s own awareness of what one “really” is can make. These are examples where a category is intuitively social but is not plausibly a perceptually individuated social position.<sup>36</sup>

Debunkers often offer the following criteria for being a social kind: that it picks out a kind that is socially constituted, or that is unified by a social property. However, if we think of social constitution (a relation of metaphysical dependence) as the defining feature of social kinds, it seems we do not sufficiently single out the kinds of particular interest to the debunker. For instance, this characterization would lead us to include kinds like *money*, *president*, *judge*, *point guard*, etc. I think the phenomenon of particular interest to the debunker is when a kind that appears natural is in fact social. However, there are plenty of socially constituted kinds that

---

<sup>36</sup> Later I will argue that they are social kinds in the preferred sense of being kinds that are featured in regularities due to joint intentional mechanisms with certain sorts of contents to be explained.

are not social in this covert sense that debunkers are interested in. Thus, this characterization appears to be too broad to capture what debunkers want to capture.

The social-positional theorist proposes that social kinds are socially constituted kinds, kinds that cannot be defined without reference to social factors. However, it is conspicuous that this proposal does not single out covert cases of construction. *Point guard, judge*, and so on will count as social kinds. But these are not cases of a social kind appearing natural, which is what the debunker wants to understand. Plausibly they are categories that are unified by a social role, but they are not plausibly kinds with the relevant type of contingency. There are not systematic regularities among linebackers or vice presidents of the sort of pressing interest to the debunking social constructionist.

To make the case that the social positional account includes irrelevant cases, consider again the social-positional theorist's suggestion that we understand a kind as distinctively social when it is *socially constituted*. To say that a kind is socially constituted is to say that it cannot be defined without reference to social factors (Haslanger, 2012: 131). It is conspicuous that this characterization of social kinds does not single out *covert* cases of social construction. On this characterization, *point guard, president, judge, money*, and so on will count as social kinds. And one might think this is exactly how it should be.<sup>37</sup> However, at the heart of the debunker's project is the search for an explanation of how a kind that is social can *appear* natural. Thus, it is undesirable that standard social positional proposals do not properly highlight what is unique to covert cases of social construction or, in other words, do not provide a direct answer to the central question posed by the debunker.

---

<sup>37</sup> See for example Khalidi (2015) for a more liberal sense of social kind.

We might not want to take on essences in general or social essences in particular for theoretical reasons, or for ethical reasons, but even if they are made perfectly innocuous essences (not understood as necessary and sufficient sets of conditions, not understood as intrinsic, and so on), social essences do not capture all and only the kinds of interest to the debunker. Because the social-positional account does not properly highlight the specific mechanisms at play in covert construction, it includes irrelevant cases and fails to include relevant ones. There seems to be the same special type of contingency at play whether the real definition of the kind is a social position or not. Further, there are kinds that plausibly are best understood as unified by a social role that do not involve the special type of contingency that we are interested in. Thus, social-positional accounts do not single out the core phenomenon, and this leads it to be both under and over inclusive.

### *Ontological Collapse*

The debunker recommends focusing on relations of metaphysical dependence rather than on social causation, in order to reveal a deeper type of contingency.<sup>38</sup> Critical realism is motivated in part by the need to capture this deeper type of contingency. However, it seems that focusing on countenancing of real social positions may inadvertently make it harder rather than easier to take seriously the very mechanisms the debunker aims to reveal. The focus on relations of metaphysical dependence between social structure and natural structures seems to make social construction just another problem of material constitution, and even to encourage *epiphenomenalism* about social phenomena. To say that it may render social phenomena epiphenomenal is to say that it renders such phenomena only *apparently* causal. Non-mentalistic

---

<sup>38</sup> See discussion of Haslanger's (2012: Ch. 3) response to Hacking's (1999) account of social construction earlier in this chapter and again in chapter two.

accounts of the phenomena of interest, e.g., those that focus on countenancing real social structures, seem to inadvertently allow the phenomena of interest to collapse into natural structures. On accounts that focus on constitutive social construction, such as Kate Ritchie's structuralist ontology of social groups, the social structures within which social kinds are thought to occupy nodes seem not sufficiently distinguished from perfectly natural structures.

Take Kate Ritchie's (2020) account of social groups. On her account social groups are *nodes within social structures that are realized by networks of relations among social animals*. Ritchie distinguishes between two different types of social group: *organized groups* and *feature groups*. In the case of organized groups there is usually an *internal* structure that is "overt and intentionally created" (Ritchie, 2020, p. 9). We may think here of sports teams, or dissertation committees, or the supreme court. In the case of feature groups, there is a structure that is *external* to the group, and the feature group occupies a *node* in that structure (Ritchie, 2020, p. 15). Often, the nodes made available by such a structure impose requirements on their occupiers (Ritchie, p. 7). Living in a socially structured environment, and having a certain property or cluster of properties, certain nodes or positions in that structure and not others are made available to one. Here Ritchie means for us to think of what we have previously been referring to as *social kinds*, e.g., races and genders. Indeed, she suggests that race and gender kinds are nodes (or occupy nodes) in social structures.

An important part of what is being claimed is that such structures are not necessitated by the laws of nature: Ritchie writes, "they depend on social factors that could (and in some cases should) be different" (Ritchie, p. 5). Examples of social factors include "social practices, habits, beliefs, intentions, arrangements, and patterns of actions" (Ritchie, p. 5). Put in technical terms, such structures are *nomologically contingent*. However, it seems to me that, on Ritchie's



account, the social structures within which social groups occupy nodes are not sufficiently distinguished from ordinary or perfectly natural structures. For instance, Ritchie compares social structure with the molecular structure of table salt. Only certain kinds of atoms can occupy the nodes or positions in the structure of table salt, just as only individuals with certain properties can occupy the nodes or positions in a social structure (Ritchie, 2020, p. 5). I worry that an account that allows for the comparison of these radically different things cannot capture the deep contingency that is the crucial insight of social constructionism. Without mention of their (causal, not just metaphysical) dependence on shared intentionality, social structures do not seem to be sufficiently distinguished from perfectly natural structures. Thus, it seems the deep contingency of the circumstances members of social kinds find themselves in is not sufficiently exposed and that social structures collapse easily into natural structures.

Ritchie and other debunkers may respond that social groups and kinds (i.e., social structures) being constitutively dependent on social factors is a major difference that does not allow them to collapse into the realm of chemical structures. If anything, it might be the consensus view that social structures are precisely not reducible to natural structures (Haslanger 2016; Barnes 2017). The strongest argument for this may be that social structural explanations are not translatable into explanations at lower levels. However, I suspect that without mention of shared intention, this picture fits too easily into a grounding framework such as is recommended by Jonathan Schaffer (2017), which precisely does allow social structures to collapse. On my view, the crucial insight of *social constructionism* is that collective attitudes partially explain the regularities that obtain among members of human kinds. The point of *metaphysical grounding* is to legitimize high-level phenomena by explaining them in terms of more fundamental things

(and so ultimately, somewhere down the line, to pin them down to the most basic physical entities). I will argue that the grounding framework renders social construction *epiphenomenal*.

Social constructionists about, for example, race or gender hold that members of these kinds share nothing but a social position, and that they occupy this position in virtue of being (regularly or for the most part) *taken* to have some feature and being treated in certain ways as a result (Haslanger, 2012). In response to the worry that the metaphysics behind social construction “remains obscure” (Mallon, 2013), Jonathan Schaffer (2017) has proposed that we understand social construction claims as grounding claims. In particular, he endorses: *to be socially constructed is to be grounded in distinctive social patterns*. However, in order to see whether “social construction as grounding” can be made to work, we have to ask about the ordering of “what grounds what” in between the mainly social (“distinctive social patterns”) and the purely physical (the most basic physical entities). If degrees of fundamentality “vary inversely with degrees of complexity” (Barnes, 2012, p. 894), surely individuals are somewhere in between the mainly social and the purely physical on the spectrum of fundamentality. So, it is natural to think that on the picture provided by the grounding theorist, social phenomena will somewhere down the line be grounded in individual phenomena. By individual phenomena I simply mean individual humans and their goings-on (their traits, their behavior, and what is in their individual heads at present); by social phenomena I mean things like collective activities, cultural practices, and so on.

As discussed, the critical realist holds that in addition to natural kinds, there are socially constructed kinds: kinds that essentially share only a social property, and yet which are no less *real* for being social. A social property is a property we have “in virtue of something about other people” (Ásta, 2018). Just as a pitch is a strike in virtue of the umpire judging it to be, and just as

an action is pious in virtue of the gods loving it, we belong to social kinds in virtue of other people regarding us in a certain way (Asta 2011;2013). Sometimes learning that a kind is social is quite surprising, as in the case of gender and race, where there are seemingly natural properties with which the social property can get confused (Haslanger, 2012; Ásta, 2018).

Although social constructionism has become an increasingly commonplace position, some have expressed puzzlement about how to fit social construction claims into a general metaphysics, and how social construction could manage to back explanation (Mallon, 2013; Schaffer 2017). In response to Mallon's worry that the metaphysics behind social construction "remains obscure," Jonathan Schaffer (2017) has suggested that in order to clarify the metaphysics behind social construction, we need to subsume it under a general metaphysical relationship and account for its apparent power to back explanation. He proposes that we understand social construction claims as grounding claims, since grounding is supposed to be a general metaphysical relationship that has the power to back explanation. In particular, he endorses: *to be socially constructed is to be grounded in distinctive social patterns.*

Grounding is supposed to be understood as the "deeper" relation of metaphysical dependence underlying the "surface" relation of supervenience (Schaffer, 2453; Kim, 1993: 167)). Where supervenience claims are thought to simply assert an *unexplainable* covariation between high level properties and their lower level base properties (Kim, 2006: 556), grounding claims are intended to assert a *generative* relation of "metaphysical causation" from the more to the less fundamental (Schaffer, 2453). That is to say, grounding is supposed to be an *explanatory* relation. If one claims that mental properties supervene on physical properties, one is simply claiming that whenever there are those physical properties, there are those mental properties. One may say this and still believe that this covariance is *in principle* unexplainable (Kim,

Jaegwon (2006). If one claims that mental properties are *grounded* in physical properties, one is making the stronger claim that the physical properties somehow explain or account for the presence of the mental properties. In summarizing the key aspects of the “inferential role” of grounding, Schaffer states: “the grounded is non-fundamental, and is generated by, dependent upon, and explicable on the basis of the grounds” (Schaffer, 2454).

Schaffer (2017) understands grounding as a primitive relation that links “a more fundamental input to a less fundamental output” (Schaffer, 2453). He motivates the grounding framework by noting “if one only lists what there is...one will have missed the further matter that *some entities are more fundamental than others*” (Schaffer, 2453). Given his examples, it seems that Elizabeth Barnes (2012) gets it right when she observes that the metaphysical notion of fundamentality often goes hand in hand with what she calls *levels ontology*. Within levels ontology there are degrees of fundamentality, which “vary inversely with degrees of complexity” (Barnes, 2012 p. 894). For example, Schaffer writes: “particles are more fundamental than chemicals, and chemicals are more fundamental than animals” (Schaffer, 2453). But the claim is not just that there are levels, it is that the fundamental entities generate the non-fundamental ones: “Once one distinguishes the more from the less fundamental, it is natural to posit a generative relation of ‘metaphysical causation’ from the more fundamental to the less fundamental. This is grounding” (Schaffer, 2453).

The point of grounding is to explain something less fundamental (for example, social construction) in terms of something more fundamental (for example, “distinctive social patterns”). It seems plausible to suggest that the most fundamental things for a physicalist will be the most basic physical entities, whatever those turn out to be. As vague as the idea of a “distinctive social pattern” might be, gesturing at it allows one to begin to see how social

construction might be pinned down to the fundamental physical entities. Schaffer asks: “what is it about social construction that allows it to back explanation? Very few relations possess the power to back explanation. Why think social construction is among them?”(Schaffer, 2452). The thought, I take it, is that we can account for social construction’s apparent power to back explanation if we can understand it in terms of ultimately physical things, the causal relationships of which can of course back explanation).

My concern is the following: if what is grounded is “non-fundamental, and is generated by, dependent upon, and explicable on the basis of the grounds,” (Schaffer, 2454) then to say that, for example, gender is grounded in social patterns, that are grounded in societies, which somewhere down the line are grounded in particles and fields, is (ultimately) to say that gender is generated by, dependent upon, and explicable on the basis of particles and fields, and this is precisely *not* what I take social constructionists to be saying. Social constructionists do not at all seem to want to say that at any given instant in time, social reality can be explained by an arrangement of physical bits. Of course, this claim of mine is only accurate if grounding is *transitive*. Here I will assume that grounding is transitive, as I understand the very point of Schaffer’s grounding framework to be to relate the non-fundamental to the fundamental. See Schaffer (2009), which assumes transitivity, and Schaffer (2012), which questions transitivity (while “retaining metaphysical structure”).

According to the social constructionist, to be raced or gendered is to be placed in a hierarchical power structure (Ásta, 2018, p. 41). Although the social constructionist might say that social structures arise from social patterns, it has been argued that the structure should be understood as distinct from those patterns:

“ Social structures, for Haslanger, are created by complex, repeated patterns of interpersonal social interactions. But they are not identical to those patterns. Their existence is both caused by and sustained by those patterns of interaction, but the structures themselves are something else. Moreover, once a structure exists, it explains the continued existence of the particular pattern, and makes that pattern harder to change” (Barnes, 2016, p. 8).

Nevertheless, Schaffer thinks that to be socially constructed is to be grounded in distinctive social patterns. Perhaps we can understand a distinctively social pattern as one that arises from shared meanings: Haslanger has stated that *gender is the social meaning of sex* and *race is the social meaning of “color”* (Haslanger, 2012). In that case, it would seem that distinctive social patterns would ontologically depend on just what is in our individual heads at this very moment (and, of course, the physical environment). This account would not seem to do justice to Barnes’ interpretation of Haslanger as a realist about social structure.

“Haslanger presents us with an intriguing idea: the joints in reality might go beyond the joints in nature” (Barnes, 2016, p. 10).

“On this view, social structures are real—as real as anything—but they are made. They aren’t joints in nature, they’re joints in the social world. We created them, and our collective social activity is responsible for their continued existence. But they’re no less real as a result” (Barnes, 2016, p. 9)

Can this be avoided by admitting diachronic building blocks? Brian Epstein (2016) writes:

“If a theory only gives us synchronic building blocks for the social facts, then that theory must be mistaken. For instance, a theory cannot be right if it takes the social world to ontologically depend only on the current beliefs, attitudes, and even current practices of community members” (Epstein, 2016, p. 154).

On Epstein's view, “many (if not most) social phenomena are built out of diachronic parts” (Epstein, 153). I do think it is more plausible, i.e., at least possible, to ground social construction diachronically as Epstein suggests. However, I will now argue that doing so (grounding, whether synchronic or diachronic) would render social construction *epiphenomenal* and so is incompatible with taking social construction claims at face value.

Jonathan Schaffer (2017) accepts that idea “that one can give sufficient explanations for all phenomena from the fundamental (e.g., from particles/fields)” (Schaffer, 2459). He elaborates:

“Perhaps the best analogy for the role of ‘fundamental entities’ in my grounding framework is that of ‘initial conditions’ in a deterministic causal model. From the causally initial conditions, the world unrolls ‘horizontally’ across time; likewise, from the fundamental conditions, the world ascends ‘vertically’ up levels” (Schaffer, 2459).

Schaffer seems to understand the two different explanatory relations that have been at issue in my discussion in this way: causation is explanatory of the way the world unrolls “horizontally” and grounding is explanatory of the way the world ascends “vertically”. He nevertheless urges that this idea is compatible with non-fundamental entities existing “fully, and in the one and only sense of ‘exist’” (Schaffer 2459) and with their having causal powers (Schaffer, 2459). However, I will now review a lesson from the British Emergentists in order to demonstrate that high-level causation on the grounding framework is simply a way of talking, i.e., that the grounding framework renders high-level phenomena epiphenomenal.

The British Emergentists proposed that there are properties that supervene on physical properties, and yet which are not reducible to them: they believed that a purely physical system that reached a certain degree of complexity could have properties not had or explained by the presence of its simpler constituents (Kim, 2006, p. 548-551). To say that such properties *supervene* on the physical properties of the system, or the properties at the “basal” level, is to say that systems with the same constituent microstructures must have the same emergent properties (Kim, 550). According to Kim, they conceded this simply because it would make no sense to hold that, e.g., a particular mental property “emerges from” a particular neural condition if the connection between the two were “irregular, haphazard, or coincidental” (Kim, 551). They

nevertheless denied that reductive explanation of emergent was possible; they did not think an explanation could in principle be given for *why* these particular properties arise from that constituent microstructure (Kim, 551).

The British Emergentists strongly believed that the emergence of mental properties makes a difference in the world at the physical level, i.e., that mental properties are causally efficacious:

“For they want to claim that the emergence of consciousness and rational thought has made a fundamental difference to the world at the physical level. It is because of our emergent mental powers that we have built cities and bridges, sent spaceships to Jupiter and Saturn, destroyed rain forests, and burned holes in the ozone layer. But can we understand how downward causation is possible?” (Kim, 558).

The British Emergentists believed in downward causation, and this was a problem for them because they also conceded that emergent properties supervene on physical ones. If one concedes that, for example, mental properties supervene on physical properties, then in order for a mental property M to bring about another mental property M\*, it seems it must do so by bringing about the underlying physical property (the “basal condition”) of M\* (namely, P\*) (Kim, 557-558). But if the mental property M emerges from physical property P, it seems that P is itself sufficient to bring about P\* (and so M\*) (Kim, 558). Emergent properties are supposed to *do* things, but their basal conditions are *already sufficient* to explain any of their putative effects: supervenience together with downward causation entails systematic overdetermination (Kim, 558).

Grounding is supposed to be the “deeper” dependency relation underlying supervenience correlations (Schaffer, 2453). If supervenience together with downward causation entails systematic overdetermination (Kim, 558), then grounding together with downward causation most certainly entails systematic overdetermination. The whole point of grounding is to allow one to talk about high-level phenomena while pinning them down firmly to low-level



phenomena. Therefore, Schaffer can say that particles and fields being sufficient to explain all other phenomena (the causal completeness of the physical) is compatible with high level causation, but it is clear that he does not mean downward causation. If the grounding framework is not compatible with downward causation, then on that framework the causal efficacy of non-fundamental entities is merely apparent, i.e., it renders high-level phenomena *epiphenomenal*.

I think that the insights of the social constructionist should lead us to reject grounding and remain agnostic about the relationship between levels of explanation.

Barnes writes, “Rather, individuals collectively interact to form a complex system. And once that system gets complex enough, properties of that system emerge which aren’t explainable simply via reference to the individuals, and which have causal influence on the behavior of those individuals” (Barnes, 2016, p. 17).

It may be objected that being dependent on social structure is a significant difference that allows for a deeper sense of contingency. I acknowledge that this could be true so long as we do not commit to asymmetrical relations of dependence as is recommended by the grounding theorist.<sup>39</sup> If we commit to the idea that social kinds are grounded in social structures that are grounded somewhere down the line in atoms and void, a genuine sense of non-inevitability is not captured. Thus, I will recommend focusing on the mechanisms responsible in the cases of interest, and a sort of agnosticism about the relationship between levels of explanation.

Capturing social construction using asymmetric relations of metaphysical dependence seems to commit one to epiphenomenalism about social phenomena. Thus, it is also counter to the aims of the social constructionist, which is to show that social forces influence the way things go. Thus, I think it is important to highlight the crucial distinction between groups and kinds. Groups explain regularities among members of social kinds. And further, to remain agnostic about the

---

<sup>39</sup> Barnes (2018) argues that ontological dependence should be understood as non-symmetric rather than asymmetric, contrary to “metaphysical orthodoxy”.

relationship between levels of explanation. Even if we say that relations of metaphysical dependence are non-symmetric, structural accounts of social construction have something missing in their overall picture, namely, the efficacy of joint intentional mechanisms. Thus, I will recommend an agnosticism about the relationship between the levels of explanation, and an account of social construction that focuses on the mechanism behind covert social construction.

## CONCLUSION

In the first part of my explanatory critique, I argued that Haslanger and other social positional theorists draw too sharp of a distinction between natural and social kinds. In the last part, I argued that other structural accounts may not make a clear enough contrast between them. My concern with Haslanger's critical realism was that it bifurcates ontology in a way that unnecessarily forces one to accept a sort of thoroughgoing realism: implying social essentialism seems to imply realism "all the way down". My concern generally with accounts that focus on structure realization, was that they seem to allow the phenomena of interest to collapse too easily into perfectly natural structures. I also argued that social positional accounts are both over- and under- inclusive in the resulting extension of *social kind*. To avoid all of these outcomes simultaneously I will suggest that instead of focusing on relationships of *metaphysical dependence* or on metaphysical explanation, we focus on giving a *mechanistic explanation* of the deeply contingent regularities of interest to debunkers, and in doing so, explaining how it is that they can *appear* natural as on my view this is what is crucial to capture with the notion of social construction.

## CHAPTER TWO: SOCIAL KIND\*: JOINT INTENTIONAL MECHANISTIC KIND

Abstract:

The goal of the debunking social constructionist is to reveal as social kinds that are widely held to be natural (or, in some cases, to reveal as more deeply social kinds that are already widely recognized to be social). The prominent approach to such debunking has been to make a case for thinking that the individuation conditions for membership in the kinds in question are in fact social (or in fact more deeply social than has previously been recognized). In this chapter, I argue that adopting the prominent approach to debunking prevents one from answering the implicit question being posed by the debunker, namely, the question of how a plausibly socially constituted kind can come to appear more natural than it in fact is. I then sketch an alternative way of understanding social kinds that enables us to answer the appearance question while remaining neutral on the nature of social kinds.

### DEBUNKING, CONSTITUTIVE SOCIAL CONSTRUCTION, AND THE APPEARANCE QUESTION

A major aim of the debunking social constructionist is to reveal as social kinds that are widely held to be natural (or, in some cases, to reveal as more *deeply* social kinds that are already widely held to be social) (Ásta, 2018: Ch. 2; Barnes, 2017; Díaz-León, 2019; Haslanger, 2012: Ch. 6; Mallon, 2008). A list of paradigmatic cases of interest includes races, genders, disabilities, sexual orientations, ethnicities, and other protected categories. This major aim is usually understood by the debunker to be in service of the larger aim of revealing oppressive social arrangements. Showing a kind to be more social than has previously been recognized is meant to show that the circumstances members of the kind find themselves in are more within human control than has previously been recognized. Thus, it is thought also to show both that such circumstances require further justification than has so far been given and that they can be changed (Ásta, 2018; Haslanger, 2012; Díaz-León, 2019).

Debunkers are often interested in how it comes about that it can seem natural, appropriate, or even morally correct for individuals with certain traits to occupy certain positions

in society and not others (Haslanger, 2012: Ch. 3). As Barnes puts it in discussing Haslanger's notion of "unmasking", "one goal of a successful theory is to explain how something we thought was a natural category 'is in fact social'" (Barnes, 2017). According to Ron Mallon, some of the most interesting putative objects of social construction "are those that are apparently natural categories of person" (Mallon, 2016, p. 2).<sup>40</sup> Similarly, Díaz-León writes that "social constructivists about human kinds such as gender, race and sexual orientation often make an additional claim, namely, that these kinds are social constructions but they are typically believed to be biological kinds (that is, people are typically wrong about the nature of these kinds)" (Díaz-León, 2019, p. 34). Debunkers often hold that justification for the circumstances that members of putatively social kinds find themselves in are not demanded due to the persistent illusion that their being in such circumstances is natural or inevitable. Thus, the goal is to reveal the cogs and belts that sustain this illusion (Ásta, 2018: 4).

In recent years, the most prominent approach to the debunking project has been to make a case for thinking that the *individuation conditions* for membership in a kind are social, that in defining the kind we must make reference to social factors, that the category in question "in fact tracks a group that occupies a certain (usually 'thick') social position rather than one defined by a set of ordinary physical or metaphysical conditions" (Haslanger, 2012, p. 132).<sup>41</sup> In every case, it is argued that belonging to the social kind in question is a matter of being *taken* to have certain

---

<sup>40</sup> He continues, "By 'apparently natural' I mean that there is a widespread tendency to treat the categories as *natural kinds*, membership in which is explained by possession of some natural (often understood as biological) property that also explains the typical features of category members" (Mallon, 2016, p. 2).

<sup>41</sup> For instance, Sally Haslanger has argued that to belong to a gender kind is to occupy a position in a hierarchical social structure as a result of being *taken* (regularly or for the most part) to have certain properties indicative of a particular role in sexual reproduction (Haslanger, 2012). She has argued similarly that to belong to a race kind is to occupy a position in a hierarchical social structure as a result of being *taken* (regularly or for the most part) to have certain properties indicative of ancestral links to a certain geographical region.

traits rather than of *having* certain traits (Ásta, 2018, p. 3).<sup>42</sup> For these are the conditions that determine which constraints and enablements we receive and so for the purposes of the debunker, may be seen as appropriately determining the membership of the kind in question. This idea is often captured with the notion of *constitutive social construction*. Constitution is a relation of *metaphysical dependence*, where one thing is said to depend for its existence on another thing. Constitutive social construction denotes a relation of metaphysical dependence on the social world or on social structure (Haslanger, 2012: Ch. 2; Ásta, 2018: 41; Barnes, 2017; Ritchie, 2020).

On the sort of view just described, social kinds are kinds that depend for their existence on social structure, and so are sometimes thought to have a social *nature*. For instance, Sally Haslanger (2012: Ch. 7) suggests that the reference of our social kind terms are positions in a social structure.<sup>43</sup> Elizabeth Barnes interprets Haslanger as suggesting that social structures are not “joints in nature” but rather “joints in the social world”. She takes this to be that same as saying that such structures (and presumably the positions within them) are as real as anything, even though they are *made* (Barnes, 2017). As Esa Díaz-León puts it “Social constructionism, then, is a version of realism about a kind that asserts that the kind is real, and puts forward a particular view about the nature of the kind, namely, that it is constituted by social factors and practices” (Díaz-León, 2019: 34). This situation raises the question: how can a kind that is

---

<sup>42</sup> As noted, the debunker might also be interested in kinds that may already be widely held to be social in some sense, such as *refugee* or *widow* (Haslanger, 2012: Ch. 3; Ásta, 2018: 36). These categories are relationally defined on any definition (e.g., take the commonsense view that to be a widow is just to have lost a husband). But the social constructionist may aim to show that, in fact, the unifying trait of the kind is a social position (e.g., that to be a widow is just to occupy a low social position or have a particular social status as a result of being *regarded* as one who has lost a husband).

<sup>43</sup> For a related but different sort of view along these lines see Ásta’s (2011; 2013; 2018) conferralist account of social kinds on which, e.g., gender is a social status (consisting in constraints and enablements) conferred on those who are *taken* to have a certain grounding property in a context (depending on the context, the grounding property may be certain sexual characteristics, or a certain domestic role, and so on).

plausibly constitutively socially constructed come to *appear* natural? I will refer to this question as *the appearance question*. I take this to be one of the central questions raised by the debunking social constructionist, but I do not think that they can satisfactorily answer it.

If we want to reveal that a kind widely held to be natural is in fact social, which I take to be the express aim of many social constructionists, we cannot accomplish this by simply asserting that the conditions for membership in the kind are social. It is the wrong sort of “explanation” for the job. When we ask how a plausibly constitutively socially constructed kind comes to *appear* natural, it is not the right kind of answer to say: this is the nature of the kind. Furthermore, simply asserting that the nature of the kind is social provides us with no tools for explaining the difference between *overt* and *covert* construction. Debunkers are interested primarily in *covert* construction: the social construction of kinds that nevertheless manage to appear natural and thus inevitable, making the demand for justification of the circumstances among members seem inappropriate (Ásta, 2018: 36). However, the analysis of social kinds as kinds with social individuation conditions fails to pick out all and only the kinds of specific interest to debunkers, and so (relatedly) fails to highlight the unique explanation required to answer the appearance question. Or so I shall argue.

In this chapter, I make a case for thinking that the appearance question is best answered with a *mechanistic* rather than a metaphysical explanation. I aim to give an account of the metaphysics of social categories but defend a form of *causal* rather than *constitutive* social construction (see Haslanger, 2012: Ch. 2; Ásta, 2018: Ch. 2).<sup>44</sup> It is not that no social kind is constitutively socially constructed, but rather that this is not the defining feature of a social kind

---

<sup>44</sup> I argue that it is a form of causal construction that does not fail where debunkers have taken Hacking’s (1999) account to, and which highlights the importance of certain mechanisms not found in Mallon’s (2016) account.

in the sense of interest to the debunker. I posit a kind of mechanism that explains the cases of primary interest to the debunker, namely, when a plausibly constitutively socially constructed kind comes to appear natural. I argue that the kinds of interest to the debunking social constructionist are not kinds of interest in virtue of *how they are individuated*, i.e., not due to sharing the same sort of “metaphysical explanation,” but rather in virtue of featuring in the same type of *mechanistic* explanation: a joint intentional mechanistic explanation. The explanation I provide here will allow for a new way of understanding what it is to be a social kind, one that I argue enables us to isolate all and only the cases of specific interest to debunkers, and to explain rather than merely label the difference between covert and overt construction, in addition to providing a straightforward answer to the appearance question.

I begin in section 2 by discussing the contrast between metaphysical and causal explanation. I then argue that something other than a metaphysical explanation is needed in order to answer the appearance question, namely, a *joint intentional mechanistic explanation* (which is a type of causal explanation). In section 3, I introduce the notion of a joint intentional mechanism, and describe the *content* of the joint intentions that are relevant to answering the appearance question. In section 4, I argue that the account of social kinds that results from this sort of explanation gets the extension of *social kind* right, and that it also allows us to explain rather than merely label the difference between overt and covert construction. In section 5, I anticipate and respond to some worries, and conclude in section 6.

## SOCIAL CONSTRUCTION AS MECHANISTIC EXPLANATION

### *Metaphysical vs. Causal Explanation*

To see the significance of the suggestion that we understand social construction as a type of mechanistic explanation (which itself is a type of *causal* explanation), it is helpful to first

contrast causal explanation with another type of explanation that many philosophers in general—and social ontologists in particular—take themselves to be properly concerned with, namely, *metaphysical explanation*. A *metaphysical* explanation asserts a relation of dependence between entities of one sort and entities of another sort, typically one being dependent on the other for its existence.<sup>45</sup> For example, the Bernini sculpture *Apollo and Daphne* (Bernini, 1625) depends for its existence on the marble which constitutes it (and perhaps some other stuff, such as Bernini’s intention circa 1625 to make a sculpture, or something about the current viewers of the sculpture). Sometimes metaphysical explanations assert a relation of dependence between *facts* rather than between entities. For instance, one might say something like “the fact that *a* is *F* grounds the fact that *a* is *G*,” (Audi, 2012), or suggest that the chemical facts ground the biological facts (Bliss, Ricki and Kelly Trogdon, 2021).

Some philosophical notions that denote relations of metaphysical dependence include, e.g., supervenience, grounding, constitution, and explication. The important point for my purposes is just that all such relations are naturally contrasted with *causal* relations. For example, Jonathan Schaffer (2017) contrasts them when he suggests we think of the grounding relation as unfolding “vertically” and indeed contrasts this with causal relations which unfold “horizontally”. Jaegwon Kim (1990) contrasts them when he notes that supervenience is exclusively a philosopher’s concept, whereas the notion of causality is ubiquitous in everyday life. Another important way of distinguishing them is that relationships of metaphysical

---

<sup>45</sup> Elizabeth Barnes (2018) has argued that the relation of ontological dependence can and should be understood as nonsymmetric (against the “metaphysical orthodoxy” that ontological dependence is asymmetric).



dependence are sometimes said to be *synchronic*, in that they can be imagined at a moment in time, whereas causal relationships must be viewed as diachronic, or occurring over time.<sup>46</sup>

Brian Epstein (2015a) finds it a crucial point of entry into social ontology to distinguish it from the inquiry into how the social world *causally* works. He suggests that social ontologists are concerned rather with investigating what the social world *is*—its building blocks or what it consists of (Epstein, 2015a, p. 151). Of course, it matters little whether my project is properly viewed as falling into the field of social ontology. I bring this suggestion to our attention only to show how common it is to view the investigation of its building blocks and their dependence or determination relations as the major aim of the philosopher concerned with the social world (if not only the metaphysician so concerned). More directly relevant to my concerns is the widespread suggestion, in particular among debunking social constructionists, that the important claim to make for the purposes of debunking projects is that what distinguishes the kinds of interest are their being *socially constituted* (Haslanger, 2012; Ásta, 2018; Barnes, 2017), their being nodes in social structures (Ritchie, 2020) or perhaps their having social grounds (Schaffer, 2017; Epstein, 2015a). Either way, the focus is on metaphysical explanation.

As we have seen, an example of a *metaphysical explanation* is one that provides the conditions for membership in a kind, or for counting as some x generally. It often provides an answer to questions that ask *in virtue of what* something belongs to a category or kind. In claiming that the kinds of interest are socially constituted (or grounded), the debunker is providing a metaphysical explanation of the kind. This contrasts with the standard notion of a

---

<sup>46</sup> That grounding is synchronic could be a point of disagreement, however. For example, Schaffer (2017) treats grounding as synchronic, and Epstein (2015b) argues that grounding can be diachronic.

*scientific explanation* that is an answer to *why it is* that some observed regularity occurs, i.e., what *causes* a given regularity among members of a particular category.<sup>47</sup>

Some may think it inappropriate for philosophers to concern themselves with causal explanations. Debunkers, quite separate from whatever it is that motivates these concerns, have argued that focusing on social causation simply cannot do justice to the role we play in sustaining oppression. That can be seen as a response to Ian Hacking's (1999) suggestion that the important or central claim of the social constructionist is that what is socially constructed is importantly *non-inevitable*. The debunking social constructionist does not however find it sufficient to emphasize non-inevitability. It is not just that, e.g., gender categories could have been conceived of differently and shaped history and individuals differently than they in fact did.<sup>48</sup> Rather, it is that to belong to, e.g., a particular gender, *just is* to belong to a particular position in a hierarchical social structure (Haslanger, 2012: Ch. 3; Barnes, 2017).<sup>49</sup> However, I will argue that there is a type of causal explanation—joint intentional mechanistic explanation—that is in fact required to do justice to the role we play in sustaining oppression. It is not my task to demonstrate the causal explanation of the regularities observed among any particular kind of interest, but instead to argue in favor of the hypothesis that figuring in a particular sort of causal

---

<sup>47</sup> In the scientific explanation literature, it is sometimes suggested that after Hume, there was an inclination to replace talk of causes with talk of regularities alone. That is of course *not* to say that a regularity is evidence of an underlying causal explanation, although it seems fair to say that it is often taken to suggest one. No one can deny the regularity, e.g., of the disproportionate incarceration of Black men in the United States. Simply noting the regularity is not enough to tell us that there is a causal explanation or what the causal explanation might be. But importantly for the debunker, a particular causal explanation is often inferred (and inferred incorrectly).

<sup>48</sup> Hacking (1999) also famously introduces and emphasizes the notion of *looping*, where members of the kind come to internalize what is said about the kind and come to emulate the kind more once labeled as belonging to it. Sally Haslanger (2012) and other debunkers have stressed that one need not be *aware* of how one is perceived or *internalize* anything to be impacted by shared conceptions and social practices.

<sup>49</sup> Related views include Ásta (2018) and Ritchie (2020).

explanation is what singles out the kinds of actual interest to the debunker (rather than having social individuation conditions).

### *Why Mechanistic Explanation?*

Oftentimes discussions of the different theories of scientific explanation begin by contrasting knowledge-that and knowledge-why: knowledge-that is merely descriptive, whereas knowledge-why is thought to be the type of knowledge that yields scientific understanding (Salmon, 1990).<sup>50</sup> To have a scientific understanding, then, is to have an explanation of why, and an explanation of why seems quite naturally to require knowledge of *causes*. While it has been argued by some philosophers of science that an explanation is simply an answer to a why question (van Fraassen, 1980), a satisfactory answer to such a question is often a *causal* one.<sup>51</sup> As we will see, the framework introduced by the new mechanists (Machamer, Darden, and Craver, 2000) reaps the benefits of causal explanation without the dire consequences that the Causal-Mechanical Model (Salmon, 1984) had for the autonomy of the special sciences (Woodward and Ross, 2021; Andersen, 2014).<sup>52</sup>

For those who find it intuitive that what we are really interested in when we ask for an explanation is a causal story, the prominent Causal-Mechanical Model developed by Wesley Salmon (1984) has a strong initial appeal. On this model, causal processes are those which are capable of transmitting marks, or changes in structure due to marking interactions, and thus of propagating a causal influence across space-time locations (Salmon, 1984).<sup>53</sup> This sort of

---

<sup>50</sup> “It is explanatory knowledge that provides scientific understanding of our world” (Salmon, 1990, p. 3).

<sup>51</sup> Of course, the idea that an explanation *need* not always involve causal claims can be traced further back to the likes of Aristotle. I am here specifically interested in relatively recent developments in the scientific explanation literature.

<sup>52</sup> It is out of the scope of this paper to defend the new mechanist’s framework. I will however argue that their framework is useful for the purposes of the debunking project.

<sup>53</sup> A later version of Salmon’s account proposes that causal processes are those which are capable of transmitting conserved quantities (such as energy or charge) (Andersen, 2014).

account captures well the starkly physical phenomenon that tends to come to mind when most people hear the word “mechanism”. However, the Causal-Mechanical Model is “notoriously” unable to allow for spatiotemporal gaps in a mechanism (Andersen, 2014), and so there has been an issue of how such an account can be made compatible with the successful explanation of the behavior of complex systems (Woodward and Ross, 2021). The point is that Salmon’s account seems not to allow that there are legitimate causal processes at play in the explanations of the behavior of complex or “higher level” systems offered by, for instance, biologists, psychologists, and economists (Woodward and Ross, 2021). If we have the intuition that there must be genuine causal processes within complex or higher-level systems, e.g., if we think we should be able to talk about the causes of the behavior of a gas, rather than the individual molecules that comprise it alone, then perhaps we should look for a way to be consistent in saying so. This is what the new mechanists (Machamer, Darden, and Craver, 2000) have attempted to do.

That is not to say that the new mechanists are the *only* ones that have done so. For example, metaphysicians like Lynne Rudder Baker (2002; 2013) worked for years on the notion of *constitution* with similar motivations in mind (as stated, I will later make my case for thinking that constitution cannot help in answering the appearance question), and very recently philosophers of science such as Lauren Ross have been working on developing ways of understanding the social structural explanations of the sort posed by Sally Haslanger as distinctly *causal* explanations (Haslanger, 2016; Ross, 2023). Ross proposes that we understand the causality of social structural explanations using the *interventionist* framework (Woodward, 2003) and by building on the—perhaps implicit—suggestion found in Haslanger’s work that social structure “can be explanatory in virtue of operating as a causal constraint” (Ross, 2023). On an interventionist framework, “to say that X is a cause of Y means that an intervention that changes

the values of X, in some background conditions B, will produce changes in the values of Y” (Ross, 2023, p. 8). So, for example, changing social structures so that resources are allocated differently will produce changes in well-being among members of certain communities. This approach shows promise for explaining the causality of certain environmental factors that shape the social kinds of interest to debunkers, such as who has access to certain resources. However, I do not think it can do full justice to the role we play in sustaining oppression or provide a complete answer to the *appearance question*. I will argue that a *joint intentional mechanistic explanation* is a part of the complete causal story of the regularities observed among social kinds.

In their pioneering article “Thinking About Mechanisms,” Machamer, Darden, and Craver (2000) lay out the groundwork for a new understanding of the term “mechanism”: the new mechanisms are to be understood as piece-by-piece descriptions of the regularities observed in the different sciences, and indeed as integral to scientific practice (Andersen, 2014). On their picture, the phenomena studied by the different sciences can be seen as having “a particular hierarchical structure comprised of nested levels” (Andersen, 2014, p. 275), and thereby providing “explanatory autonomy to the so-called special sciences, while still grounding their objects of investigation firmly in the physical world” (Andersen, 2014, p. 274) and “without having to reduce or eliminate higher levels or treat them as merely apparent or epiphenomenal” (Andersen, 2014, p. 275). In other words, even if the mechanisms of complex systems are ultimately comprised by lower-level mechanisms, they are not thereby reduced to these lower-level mechanisms: their view is that the higher-level sciences provide genuine explanations that “bottom out” without being ultimately grounded in or reduced to the “sub-sub-submechanisms” of physics (Andersen, 2014, p. 276). Which is to say that the focus is just on the mechanisms at play at the relevant level of description and that those descriptions are taken at face value, not

seen merely as convenient ways of speaking. Their methodological approach allows, and perhaps can be seen even as encouraging, agnosticism regarding ultimate ontology (or an ultimate causal structure of the universe, understood as independent of scientific practice) (Andersen, 2014, p. 279; Machamer, Darden, and Craver, 2000, p. 23).

According to the new mechanists, what is often taken as a satisfactory explanation in science is the discovery and description of a mechanism. Their focus is on molecular biology and neurobiology, although they suspect that their analysis is “applicable to many other sciences, and maybe even to cognitive or social mechanisms” (Machamer, Darden, and Craver, 2000, p. 2). On their view, to describe a mechanism for a phenomenon is to explain how that phenomenon was produced (Machamer, Darden, and Craver, 2000, p. 3). Again, it matters little if the explanation we provide is considered a paradigmatic scientific explanation, or a scientific explanation period. This, at least, is not a motivating concern of mine. I am interested in mechanistic explanation because it seems that if we want an answer to the appearance question—the question of how a kind that is plausibly socially constituted can *appear* natural—then we need to explain how that phenomenon was produced. As noted above, on the analysis of the new mechanist, to satisfactorily explain how a phenomenon was produced in the sciences is often to describe a *mechanism*. On the analysis of the new mechanist, describing a mechanism involves decomposing the phenomenon of interest into *entities* and *activities*. To provide a mechanistic explanation is to provide a description of the entities and activities that bring about some regularity. Entities act as causes when they engage in productive activities, and these productive activities bring about the non-accidental regularities that are observed and which we are interested in explaining (Machamer, Darden, and Craver, 2000, p. 6-8)

The new mechanists give the example of chemical neurotransmission and of DNA replication. In the mechanism of chemical neurotransmission, presynaptic neurons transmit signals to post-synaptic neurons. This occurs when presynaptic neurons release neurotransmitter molecules that depolarize the post-synaptic cell by diffusing across the synaptic cleft and binding to receptors (Machamer, Darden, and Craver, 2000, p. 3). In the mechanism of DNA replication, the unwinding DNA double helix exposes slightly charged bases that complementary bases then bond to, eventually producing two duplicate helices (Machamer, Darden, and Craver, 2000, p. 3). Just focusing on the fact that there is a helix or a charged base here and that partially constitutes the phenomenon of interest does not yet provide an explanation of how DNA is replicated. Similarly, just focusing on the fact that there is a social structure and that it partially or wholly constitutes gender does not yet provide an explanation of how it ever appeared to us that gender is natural or biologically determined.

If we want to reveal that a kind widely held to be natural is in fact social, we cannot accomplish this by simply asserting that the conditions for membership in the kind are social. It is the wrong sort of “explanation” for the job. When we ask how a plausibly socially constituted kind comes to *appear* natural, it is not the right kind of answer to say: this is the nature of the kind. Providing the metaphysical nature of the kind does not answer the question that I am posing, and which many social constructionists are interested in answering. Why do we need a non-metaphysical explanation for the debunker’s question? It is a little bit like asking, “how did it come about that this vase has monetary value?” and giving the “explanation” that the vase is constituted by glass and the intentions of the artist. The point is that this is an incomplete answer to the question. In the cases of interest to debunking social constructionists, the metaphysical

explanation is not enough to answer the central question we have about them. What is missing is the causal story that is distinctly social.

The appearance question requires a *mechanistic* explanation, not a metaphysical one. We want to know how a certain phenomenon was produced: how it came to appear that gender is natural or biologically determined, say. So, we might ask: what mechanism explains the regularities we observe among individuals with a certain sex assignment, and increase the inductive potential of having a certain sex assignment? I will suggest that when the mechanistic explanation involves joint intentionality—in which a certain sort of awareness of what others will do informs what one does—then it is a *joint intentional mechanistic explanation*. I will make a case for thinking that when a kind figures prominently in joint intentional mechanistic explanations, it is a social kind. I will make a case for thinking that when a kind figures in joint intentional mechanistic explanations with a particular type of content to be described, namely, *covertly normative representation*, it is a social kind in the sense of primary interest to the debunker.

## JOINT INTENTIONAL MECHANISTIC EXPLANATION

### *Joint Intentional Mechanisms*

There is a strong intuitive connection between the *inductive potency* of a category and the aptness of understanding it as a natural kind. It has been argued that the concept of a natural kind is essential to the inductive component of science (Bird, 2018; Mill 1843; Quine, 1969). As Alexander Bird has put it, “our success in inducing predictively confirmed generalizations—protons attract electrons, potassium is more reactive than sodium, members of the family Felidae are carnivorous, and so forth—allows us to infer the genuineness of our natural categories” (Bird, 2018, p. 3). This line of thought, that reliable predictions among category members lends



plausibility to the naturalness of that category, leads to the idea that a *social* kind is inductively potent due to something *other* than its inherent traits. I propose that many of the kinds of interest to the debunker are those that would merely be disjunctive property clusters—or else share intrinsic properties that are not very explanatory in and of themselves—and so would lack inductive potency were it not for the existence of a *joint intentional mechanism*. I will argue that joint commitments with certain sorts of contents are the mechanism that explains how social kinds can appear natural.

The common ground shared by those who pioneered the work on collective intentionality is that a shared intention does not allow that each of those who are said (collectively) to believe something, or have some other (collective) psychological property, merely have a token of the relevant belief, assumption, attitude, etc. Rather, the notion requires precisely that there is something more than a mere summation of individual acceptances at play (Bratman, 1993; Searle, 1990; Gilbert, 2009). Shared intention is something other than a summation of individual intentions. However, views of collective intention vary in how individualist or singularist they tend to be in how they get spelled out. Examples of more individualist views are those that recommend understanding joint intention in terms of *interlocking individual intentions* (Bratman, 1999) or in terms of collection of individuals that share a *we-intention* (Tuomela & Miller 1988; Searle, 1995).<sup>54</sup> Margaret Gilbert (2009) offers criticisms of singularist views and presents an alternative that invokes her notion *joint commitment*, which gives rise to rights and obligations. I think that normativity is an important feature in answering the appearance question and I will therefore adopt and develop a Gilbertian version of my view here. I do however think that other

---

<sup>54</sup> See discussion of collective intentionality in the entry by Schweikard, David P. and Hans Bernhard Schmid in the *Stanford Encyclopedia of Philosophy*.

viable explanations that adopt different views of collective intentionality might also be developed. A version that utilizes the notion of *convention* may also be tenable (Lewis, 1969; O'Connor, 2019).<sup>55</sup>

The term *plural subject* is used by Margaret Gilbert (1989; 2013) to refer to a given set of persons who are *jointly committed* in some way (Gilbert, 2013: Introduction). Roughly, for a given set of persons to be jointly committed in some way is for those persons to have agreed (either explicitly or implicitly) to act as a body that, e.g., believes that P.<sup>56</sup> The given set of persons can be jointly committed to a rule, norm, assumption, goal, etc. instead of or in addition to beliefs. The agreement to act as a body that, e.g., believes that P, involves the standing, and indeed the obligation to “rebuke” those who do not act in accordance with the joint commitments of the group or plural subject.<sup>57</sup> Importantly for our purposes, such agreements (commitments) can be implicit or explicit, and such subjects can be small or large scale (Gilbert, 2006: Ch. 8.3). Just as there could be an explicit exchange of agreements between two people, there could be an implicit agreement between two people to walk together for instance. Just as there could be an explicit agreement among those belonging to a society to obey certain rules, there could be an implicit agreement among those belonging to a certain culture to that effect.

That a joint commitment need not be explicitly agreed upon in order to have normative force can be seen in the example of two individuals walking together. We can imagine two individuals walking with one another from one class to another: maybe they never explicitly agreed to walk from one class they have together to the next, but they simply got in the habit of doing so. Intuitively, despite this lack of explicit agreement, one would be within their rights to

---

<sup>55</sup> Note that Gilbert (1989; 2018) has offered a joint commitment account of convention.

<sup>56</sup> For a much more detailed explanation see Gilbert, 2013: Introduction; Gilbert, 2023: Introduction.

<sup>57</sup> “obligation” in a sense Gilbert sees as correlative with the standing to demand.

rebuke the other if, during their usual walk from one class to another, the other darted off in the other direction suddenly and without an explanation or request to be excused from the implicit agreement that they walk together.<sup>58</sup>

As said, large scale implicit agreements are also conceivable on a Gilbertian view, and this ought to be of particular interest to the debunker. For example, if our society forms an implicit agreement that gender presentation should reflect sex assignment, then members of the society will have the standing to rebuke those who fail to comply. Such rebukes, if harsh enough, may then serve to explain why it is that in fact, those with a certain sex assignment tend to express gender in semi uniform and predictable ways. In this way, we can see how plural subjects might be the partial *cause* of a given regularity, and so also how they might plausibly feature in a *mechanistic explanation* of the regularities that obtain among collections of individuals with a cluster of perfectly ordinary (but socially significant) properties.

For our purposes, it is also important to note that to say that a collection of individuals is jointly committed to acting in accordance with a rule is precisely *not* to say that each individual in that collection is *personally* so committed. Whether all the individual members of a society *personally* accept that gender presentation should reflect sex assigned at birth or not, the fear of rebuke for not complying, not enforcing, or for trying to prevent enforcement, is apt to play a role in explaining the robustness of the regularity that gender expression tends to match the expectations for the sex one is assigned a birth. So, for example, one may not be *personally* committed to the belief that, e.g., a wife should cook and clean for their husband, and yet, as a member of a community that harbors this belief, in the sense that members have agreed to act in

---

<sup>58</sup> See Gilbert 1990

ways that express it, one may feel compelled to act in ways that comply with this belief and brace for a rebuke if they fail to do so.

Just noticing this potential mismatch in plural subject-level and individual-level intentions seems to suggest a way of revealing the source of the entrenchment of the covert constructions of particular interest to debunking social constructionists. On such a view, it is possible to say that one is in a certain sense obligated to act in accordance with a belief that one may not personally accept. Because there is at least an implicit agreement that one will act in accordance with a shared intention, other parties to the intention have the standing to demand ones acting in accordance with the content of that intention, or of rebuking one should they fail to do so.<sup>59</sup>

The debunker wants to reveal that, sometimes to our surprise, human categories are explanatory in virtue of something other than the property or properties that we think we are picking out with the category. They thus tend to emphasize that sometimes what we are in fact picking out is a shared social position, that this is the *operative concept* (Haslanger, 2012: 92). Because being regarded as belonging to this position has consequences for what we may or may not do, or have access to, it is also emphasized that such positions are *real*. But what matters for revealing that human categories are explanatory in virtue of something other than the properties we think we are picking out, is saying what exactly *does* explain regularities among category members. This does not actually seem to require complete agreement about the conditions for membership in such kinds. My hypothesis is that—whatever the nature of such kinds—an explanatory role is played by our commitment to certain representations of them, a role that is

---

<sup>59</sup> Gilbert argues at length for this relation (see, for example, Gilbert, 2018).

underemphasized on the currently prominent view. I propose that there is a social kind, rather, insofar as members of a kind are subject to certain norms that enable us to make reliable generalizations about members that we would not be able to make were it not for the presence of the joint commitment and the correlative normative expectations.<sup>60</sup>

### *The Content of the Shared Intention*

I propose that we can use Gilbertian plural subjects of a particular sort to give an analysis of the subjects of a certain sort of *covertly normative representation* and that we can then see how these relate to paradigmatically social kinds in the sense of interest to the debunker. I have suggested that the activities such plural subjects engage in constitute a mechanism that can give us an answer the appearance question.<sup>61</sup> If I want to suggest that we understand the sort of social kind of interest to debunkers in relation to plural subjects, then it will be asked just what exactly it is that the individuals that make up such entities are jointly committed to acting in accordance with. The relevant plural subjects are entities that engage in joint intentional activities that are productive of regularities among collections that would otherwise lack that same extent of inductive potency. I will suggest that one fruitful way of singling out the relevant plural subjects is to focus on those that jointly accept what I call a *covertly normative representation* of a kind.

What I have in mind is something akin to a *dual character concept* as discussed by Knobe, Prasada, & Newman (2013). Knobe et al point to five experiments they take to provide

---

<sup>60</sup> These norms might often be “negative”. For example, members of a category may be excluded from the benefits of a practice, as in a case where homosexual couples are not permitted to adopt children or marry due to there being a joint commitment to a belief that marriage or parenting is only appropriate to heterosexuals.

<sup>61</sup> We might think of plural subjects as the subjects of conceptual schema and ideology more generally. The plural subjects of particular interest here are those that are party to a joint commitment with covertly normative representations as part of their content.

evidence for a class of *dual character concepts*. They describe the notion of a dual character concept in the following way:

“Dual character concepts characterize their members in terms of both (a) a set of concrete features and (b) the abstract values that these features serve to realize. As such, these concepts provide two bases for evaluating category members and two different criteria for category membership” (Knobe, Prasada, & Newman, 2013, p. 242).

One place we can quickly see the usefulness of this idea for the purposes of the debunker is in Sarah-Jane Leslie’s (2015) paper, “‘Hillary Clinton is the Only Man in the Obama Administration’: Dual Character Concepts, Generics, and Gender.”. The idea she explores is that, if we can quickly understand the statement that Hillary Clinton (who we know to be a woman, under one important set of criteria) is a man (“the only man in the Obama administration,” as the headline she references states)—we can infer the existence of some sort of understanding or agreement that there is *another* set of criteria for being a man at play here that involves manifesting certain traits, perhaps in this case those traits might be competence, strength, and effectiveness. That set of criteria is one that Hillary meets apparently better than any of the so-called men in the Obama administration, according to the author of the headline Leslie references. I will not take on the notion of a dual character concept, because I do not think it would be correct to say that in most or perhaps any of the cases of interest to the debunker that there are two distinct delineated criteria for category membership. However, there seems often to be more than one base for evaluating category members.

In the cases I am interested in, the descriptive component of the representation in question is a cluster of plausibly explanatory properties which turns out not to be explanatory, in and of itself. Instead, the way the evaluative component of the representation is acted on in a semi-uniform fashion by a plural subject contributes significantly to the explanatoriness of the

kind represented. For example, in the case of gender, individuals may take note of certain plausibly explanatory characteristics of a person, say the biological features of a person, or their apparent sex assignment, and on some level take these very features to explain why it is that a commonly made generalization holds true, when in fact the expectation itself contributes to the reliability of the inference.

Take cases of the form, “ladies cross their legs” or “boys don’t cry”. These are statements that have the same form as statements that are taken to implicitly provide an explanation for the very generalization stated, e.g., “Mosquitos lay eggs” or “water boils at 212 degrees Fahrenheit”. It is something in the nature of mosquitos that explains why it is that they lay eggs; it is something in the nature of water that explains why it boils at 212 degrees Fahrenheit. But upon reflection, at least, we know that it is not in the nature of ladies to cross their legs, and it is not in the nature of boys never to cry. Nevertheless, these generalizations turn out to be relatively reliable. Insofar as the generalization *does* in fact hold true, it is due to the fact that when we say things like “ladies cross their legs” or “boys don’t cry” we are enforcing an implicit agreement that those who meet one set of criteria for being a lady or for being a boy—usually traits taken to be intrinsic to individuals, such as biological properties, but they need not be—ought *also* to meet these other criteria for being an exemplar of the category (manifesting certain ideal traits, e.g., being modest, being tough). The most salient examples may be so-called normative generics, but they need not be.<sup>62</sup> For example, we may be able to make the generalization that female philosophy graduate students contribute less to seminar discussions.<sup>63</sup> We might not be

---

<sup>62</sup> See Leslie and Lerner (2022) for a discussion of normative generics.

<sup>63</sup> As Ritchie (2019) argues, some social generics may be required for describing and ameliorating structural injustices.

jointly committed to acting in accordance with this belief but nevertheless be committed to representing femininity in such a way that contributes to the truth of this generalization.

On the prominent view, social kinds are kinds that are socially constituted. *Woman* is a social kind on that view because to be a woman is just to occupy a subordinate position in a hierarchical social structure (Haslanger, 2012: Ch. 7). On my view, *woman* is a social kind because regularities among women are explained in part by joint intentional mechanisms that have a particular sort of normative content. The regularities observed among women would be less robust, the category woman would less reliably figure in true generalizations, would be a less inductively potent category, if not for our joint intentional acting in a semi regular semi uniform fashion on the traits shared by women. On my view, however, this is also true of the kind *female*. For the sake of argument, we could say that to be female is just to have a cluster of biological properties. But we could also say that to be female is just to have a certain government assigned sex status. It does not matter, on my view, whether the individuation conditions for membership in the kind are social or nonsocial. What matters is what explains the regularities observed among members of the kind. Maybe as a society we are jointly committed to the belief that *females* ought to be quiet when males are speaking. This will shape regularities among those assigned female at birth. Maybe as a society we are jointly committed to the belief that *women* ought to dress femininely. This will shape regularities among those who identify as women.

## EXPLANATORY APPEAL AND ANSWERING THE APPEARANCE QUESTION

### *Answer to the Appearance Question*

The regular citing and enforcing of such implicit agreements then serves to make the statement or generalization more stable or reliable. It also serves to illustrate how joint intentional mechanisms might enable us to answer the appearance question. If the people in a



culture express readiness to act in accordance with a representation that is covertly normative, then even if there were nobody who personally accepted the implications of that covertly normative representation, the behavior of all involved would be informed by it. Fear of rebuke shapes behavior without explicit enforcement, and so contributes to the appearance that the shared behaviors result from natural properties shared by kind members. No one need believe what they act in accordance with, or act on their own behalf, and yet the coordination of the parties to the joint commitment will produce regularities among those who share descriptive components of the represented kind. Because none of this is made explicit or played out due to personal beliefs, these pervasive patterns do not *seem* like the result of anything we have put into place. If no one is explicitly forced to comply with the relevant norms or to enforce them, then the best explanation of the regularities appears to be the features that members of the kind have in common. However, if we acknowledge that in fact joint intentional mechanisms make the generalizations more reliable, then we can explain the inductive potency of the kind without implying that intrinsic properties are responsible.

#### *Other Causal Accounts and the Importance of Joint Intentionality*

It may be noticed that it is the notion of a *covertly normative representation* which seems to be the part of the model on offer here that explains *covert* construction, as I have set out to do. Indeed, *covertly normative representation* may appear to be the crucial piece for explaining how a socially constituted kind can come to appear natural. Noticing this may lead one to believe that the notion of a joint intentional mechanism is superfluous or only of secondary importance. However, the way it is working here, covertly normative representations are a type of joint intentional mechanism, and it is the joint commitment to the representation that explains the *uniform* regularities of interest to debunkers. The normative expectations that arise from joint

intention provide an alternative way of explaining the robust regularities observed among members of the kind. Behavior is perpetually shaped by normative expectations that we recognize others as having a certain sort of standing to enforce. In lieu of that explanation, there appears to be a human tendency to explain those regularities by suggesting that the features shared by kind members constitute a natural kind. The covert normativity of the representation partially explains how it can go unnoticed that norms are being enforced, the joint intention explains the uniform treatment and so the robustness of the relevant regularities.

Thus, the proposal can be distinguished from Hacking's explanation involving *looping* and so can avoid the debunker's critique of that sort of causal account. As noted, in Hacking's explanation the crucial thing causally seems to be that the kind members *internalize* something. For that reason, debunkers have rightly argued that it does not get at the systematicity of the oppressive social forces debunkers want to uncover, and wrongly suggests that one needs to be aware of such norms and constraints in order to be impacted by them. Whereas Hacking focuses on the way that representations become internalized among members of a category, making the descriptions and representations come to fit, I say that the sort of explanation needed is a *joint intentional mechanistic explanation*, an explanation focused on how collective actions contribute to the appearance of naturalness. These joint intentions have covertly normative representations as part of their content. For example, the representation *woman* could plausibly be said to have both a descriptive and a normative component. The descriptive component could involve bodily features or presentation, or other features depending on the context or culture. The normative component that is acted on may include norms and roles considered appropriate to those who meet the descriptive component.

The view can also be distinguished from another prominent causal account developed by Ron Mallon (2016). While Mallon discusses representations and how they can be causal, outside just the internalization of looping, as far as I know he does not discuss any role for *collective intention* in particular. As suggested, I think that Lewisian conventions, which are discussed by Mallon, could be understood in social mechanistic terms and so perhaps could accomplish the same thing as plural subjects do on my account. However, I think it is important to the goal of the debunking social constructionist that we reveal the causal influence of collective intention specifically, and do so explicitly, or else we may fall into the same problem the debunkers noticed with Hacking's view. It is not just that these categories could have been different. It is that the kind is explanatory in part due to the acceptance of certain norms that we are compelled to *actively maintain even if we do not personally accept them*.

The explanation offered is causal, but it is different than other causal accounts in focusing specifically on joint intentional mechanisms which explain the robust uniformity of the regularities of interest, and which do not collapse into more fundamental mechanisms (because joint commitments cannot be understood in terms of personal commitments). This is the difference and why the proposed explanation is more successful in exposing the deeper sort of contingency of interest to debunkers than are other causal accounts.

#### *A New Understanding of a Social Kind*

The account on offer allows for a new understanding of what it is to be a social kind. One that enables us to conceive of social kinds that have genuinely shared properties, that are social in virtue not of sharing a social *property* (although they may well share such a property) but rather in virtue of featuring in distinctly social *regularities*. It also allows us to answer a question that I think we should be interested in and able to coherently ask: can a kind that is wholly

natural in membership conditions, figure in social regularities? Can a plausibly natural kind come to appear more *deeply* natural than it in fact is?

Imagine a plausibly biological kind that is social in the sense of primary interest to the debunker. Suppose there is a kind that has fully natural individuation conditions, say, a kind for which there is a specific genetic marker. Now imagine that people with that genetic marker are excluded from receiving certain privileges in our society and that having that genetic marker is stigmatized. Now this kind comes to feature in social regularities. This is precisely the sort of thing social constructionists are interested in and want to explain. By definition, it is not going to be a *social* kind if one is only interested in giving metaphysical explanations. That is because the kind we are picking out is unified simply by the genetic marker. One could *choose* to build the relevant sort of social regularities into the membership conditions for the kind, as the debunker has tended to do, but there may be pragmatic considerations in favor of not doing so. For instance, it might be less divisive, it might be that doctors or scientists have a need to refer to the individuals with the physical condition, it might be that people who belong to the kind view themselves as really sharing a physical condition, and so on. In the case of having a specific genetic marker, it seems the metaphysical conditions for the kind are simply genetic, but it seems it could plausibly be considered a social kind in the sense of featuring in social regularities. Then do we really have to say that to belong to the kind that seems to be unified by that genetic marker is *really* just to have a social status?

It is a strange consequence of the prominent approach to debunking that if we want to acknowledge a kind as social in the relevant sense then we have to say that the metaphysical conditions for membership are really social even in the case of a genetic marker. On the account that I want to give, I do *not* have to say that. We can instead shift our attention away from the

nature of the kind and toward the sort of regularities in which the kind is featured. Building the social regularities into the nature of the kind does not seem appropriate in all cases. What is most interesting and what matters most, for the purposes of debunking at least, is how the regularities are being explained rather than what the nature of the kind is.

The proposal on offer is explanatorily attractive in that often there is a trait that seems like the inherent kind of trait that could potentially be explanatory in and of itself. My proposal allows us to explain how it comes to be that kinds that share such a feature appear more natural, that is, become more explanatory, than they in fact are. Our regular acting on the trait in systematic ways produces robust regularities among those that share the trait and that would otherwise not occur or be as robust. The joint intention is what produces the systematic behaviors and so explains the uniform regularities. The account allows us to make sense of kinds that are intuitively social in the relevant sense and yet do not have social constitutive conditions, as in the case of the kind unified by a genetic marker. I want to explain the regularities the kinds of interest figure in and the inductive potential they have that leads us to view them as more deeply natural than they in fact are. The generalizations we can make about members of these kinds would not be as reliable, the regularities they figure in would not occur or be as robust, if it were not for our regular acting in accordance with a particular sort of representation.

#### *Explaining the Distinction between Overt and Covert Social Construction*

The prominent approach is to say that social kinds are those that are socially constituted. This not only prevents us from including some intuitively social kinds that have nonsocial individuation conditions, it also does not give us an explanation of the distinction between overt and covert construction, and so seems overinclusive. Are the kinds *pointguard* and *president* social kinds in the same sense as *woman* is? Rather than simply saying that some kinds are

socially constituted, and in some cases, it is covert and in other cases it is overt, the proposed view provides an explanation of covert construction: why these robust regularities seem natural, and where they actually get their inductive potential from. Thus, the view can explain the difference between these two types of social kinds rather than merely label the difference.

I spell this out in the following way: rather than understanding social kinds as kinds that are socially constituted, which captures kinds that are intuitively not the paradigmatic sort of interest to debunkers, and leaves out kinds that intuitively are, we can say that the social kinds of interest to debunkers are those that feature in a certain sort of regularity. A regularity explained by commitment to a covertly normative representation. This not only allows us to pick out the kinds that intuitively are paradigmatic examples of the social kinds of interest to debunkers, it allows us to answer the appearance question. Debunkers are interested in kinds that appear natural even though they are plausibly socially constructed or constituted. Joint commitment to a covertly normative representation of the kind explains the robust regularities observed among members of the kind while answering how they appear natural. They appear natural because no one needs to explicitly agree to acting in accordance with the representation and no one need personally accept that the behaviors that are in accordance with the representation are natural, desirable, or good.

Joint intentional mechanisms are useful for capturing social construction because they explain how socially constituted kinds can figure in uniform regularities such that they appear natural. Covertly normative representations are particularly useful for singling out *covert* social constructions because their existence demonstrates that the content that is salient is not necessarily the content is acted on. There is not a different *mechanism* but there is a different type of representation, in cases of overt construction. In cases of overt construction, the

representation is overtly normative. For example, when someone is represented as a doctor, we have explicit normative expectations for their behavior and our representation of the kind is of a *role*. In cases of *covert* construction, the representation is covertly normative. For example, when someone is represented as a man, we might represent our expectations of them as a prediction based on nature, since the representation of the kind is descriptive and includes seemingly natural properties. But it is plainly false that, e.g., not ever crying is a prediction based on the nature of individuals with whatever relevant cluster of biological traits. Thus, we arrive at the happy consequence that all social kinds can be analyzed in one way, and yet the distinction between covert and overt construction is still easier to make. We can actually explain the difference rather than merely label the difference. That we are interested in covert construction reveals that what we are after is an answer to the appearance question. I have attempted to answer the appearance question and in doing so single out the kinds of interest to debunkers.

#### WORRY CASES AND ANTICIPATING OBJECTIONS

Here I will consider a couple of initial objections to the view on offer. One worry is that the outcome will be that kinds that are too *thinly* social to be of interest to the debunker will count as social, and that this is the wrong result. I will call this the *tiger objection*. Another worry is that the view of social construction provided is too cognitive and so does not well capture the role of material practices and conditions. I will call this the *too cognitive objection*.

##### *Tiger Objection*

Does this account allow in cases that we want to exclude, e.g., cases that are too *thinly* social to be of any particular interest to the debunker? If all it is to be a social kind in my proposed sense is for the kind to be shaped by the enforcement of normative expectations, then it seems lots of kinds will be social kinds. Take the *tiger objection*: tigers are individuated,

presumably, by a cluster of biological properties. But it seems plausible to hold that it is true of tigers that, to some extent, they are shaped by the enforcement of normative expectations. To make the worry more vivid, we could imagine a world where all tigers lived in enclosures, were given only one particular type of processed food, and where people were punished for acting outside of these expectations for how to treat tigers. There would then be many regularities among tigers that were shaped by the enforcement of norms for how to treat creatures that share that cluster of biological properties. The kind would be more inductively potent than it would otherwise be, were it not for the enforcement of normative expectations for the treatment of tigers. However, this kind seems to be too thinly social to be of interest to the kind of theorist I set out to address.

I accept that under such conditions the kind *tiger* would count as a social on my view, but it would not fall into the subset of social kinds of particular interest to me and to the type of theorist I set out to address, namely, debunkers. The subset of social kinds of interest to debunkers are distinguished by the type of *content* in the shared intention being acted on, and thus by a type of cause. It may be crucial to covert cases of construction that the content involves a normative dimension for the kind itself. In the tiger objection, there is a norm for *how to treat tigers*, but there is not really a normative dimension to our *representation of the kind tiger*. Does the account cover the cases that we intuitively want to capture but that are indeed plausibly socially constituted or individuated by “thick” social roles? I do not want to suggest that there could be no such thing as a social kind in my sense that is entirely socially constituted. Take a case like *widow*. The kind *widow* could pick out all and only individuals who had a husband that died, or the kind *widow* could pick out all and only individuals who are (regularly or for the most part) *taken* to have lost a husband, and who occupy a social role that is disadvantaged as a result.



Either kind could be social insofar as it is featured in regularities that are explained in part by joint intentions with covertly normative representations as part of their content.

### *Too Cognitive Objection*

Debunkers are often interested in the material conditions and practices that explain regularities among members of human kinds. Imagine a scenario where a certain political activity always takes place on the second floor of a building that has no elevators. In that case, it seems that it is a feature of the environment that explains why people with certain disabilities will not be able to take part in said political activity, such as voting.<sup>64</sup> Sally Haslanger (2016; 2020) makes a similar point when discussing the material realities involved in social regularities, such as a bus line being down and who is impacted by this sort of occurrence. One might think the view on offer here is not well suited to accommodate such cases, or explain what makes the resulting regularities social. However, might not the joint acceptance of a representation still be playing a role in explaining the social regularity in these kinds of cases?

This sort of example leads me to think that in some cases the relevant content or representation is not of the kind in question *being a particular way* but rather the members of the kind being *left out* of a representation of a broader kind that is understood as properly being another way. For instance, if we create a building for political engagement and do not make it accessible to those with certain disabilities, then perhaps it could be argued that the representation that plays a role here is about the kind *person* and what is proper to the appearance and behavior of persons, namely, that they are able bodied. I think that social structural explanations could play a complementary role in a complete explanation of category-

---

<sup>64</sup> This example comes from Ásta.

based oppression. For instance, Lauren Ross' (2023) proposal is a step toward understanding the causality of social structure. However, I think social mechanisms are also a crucial and underemphasized part of that causal story. Representations cannot do everything, but they definitely do something, and that something is currently underemphasized in the prominent approach.

Considering the role representations play when it comes to practices, such as voting practices taking place on the second floor without elevators, brings to our attention the way that sometimes the content of a shared intention that explains regularities among members of a kind is “negative” in the sense that the regularities are a result of the members of the kind being excluded from a broader kind that has a normative dimension not met by members of the particular kind of interest which is a subset of the broader kind.<sup>65</sup> Considering the role representations play when it comes to regularities that seem to result from features of the environment or from material practices, we can see the way in which our representations might *inform* our practices, including environment shaping practices that impact those who fail to meet the normative dimension of a jointly accepted representation.

## CONCLUSION

In this chapter I argued that the appearance question—the question of how a plausibly socially constituted kind can come to appear natural—is one of the central questions raised by the debunking social constructionist, and that they cannot satisfactorily answer it using the prominent approach. If we want to reveal that a kind widely held to be natural is in fact social, which is the express aim of many social constructionists, we cannot accomplish this by simply

---

<sup>65</sup> For example, homosexuals are sometimes excluded from the kind that is represented as suited to marriage, and this exclusion explains certain regularities observed among homosexuals.

asserting that the conditions for membership in the kind are social. It is the wrong sort of “explanation” for the job. When we ask how a plausibly socially constituted kind comes to *appear* natural, it is not the right kind of answer to say: this is the nature of the kind. I have made a case for thinking that our aims as debunkers are better served if we instead shift our attention away from the nature of the kind and toward the sort of regularities in which the kind is featured. The generalizations we can make about members of these kinds would not be as reliable, the regularities they figure in would not occur or be as robust, if it were not for our regular acting in accordance with a particular sort of representation, i.e., without joint commitments to covertly normative representations. Because the sort of regularities the kind is featured in is what matters, we can answer the appearance question while remaining neutral on the nature of social kinds. As we saw in the environmental cases, representations cannot do everything. However, they definitely do something, and that something is currently underemphasized in the prominent approach.

## CHAPTER THREE: THE VIABILITY OF CRITICAL ANTI-REALISM

Abstract:

Some debunking social constructionists about human categories such as race and gender have argued that *metaphysical realism* is an important component of the debunking project, and suggested the adoption of *critical realism* (Haslanger, 2012: Ch. 6). Critical realism is the view that kinds unified by a social property or position are nevertheless real kinds, that social kinds are no less *real* for being social (Barnes, 2017). The critical realist holds further that there are kinds that are determined completely mind-independently, and that retaining this contrast is essential to social change. In this chapter, I make a case for the viability of what I term critical *anti-realism* by comparing my joint intentional mechanistic kind view with Ásta's (2018) conferralist account of social kinds. I argue that while the conferralist account is a social positional view of social kinds and the joint intentional mechanistic kind view is not, they are *both* viable forms of critical anti-realism, i.e., forms of metaphysical anti-realism that meet the aims of the debunking project. I then give reasons for thinking that the joint intentional mechanistic kind view has benefits that outweigh those of the conferralist account by examining the case of sexual orientation. Lastly, I discuss how the proposed view avoids the problems of chapter one and respond to some possible objections and worries.

### INTRODUCTION

In the last chapter I recommended that by focusing on the mechanisms that explain the regularities among members of putatively social kinds, we can remain neutral on the *nature* of social kinds while nevertheless capturing that they feature in distinctly social regularities. The primary aim of my project has been to provide a straightforward answer to what I have called *the appearance question*, the question of how a plausibly constitutively socially constructed kind can appear natural. A secondary motivation for the joint intentional mechanistic kind view was to illustrate the viability of critical anti-realism, i.e., that metaphysical realism is not required to meet the aims of the debunking project. One might think that Ásta's (2018) *conferralist* account of the metaphysics of social categories already demonstrates the viability of a sort of critical anti-realism. Here I will further demonstrate the viability of critical anti-realism by showing that the joint intentional mechanistic kind view is another form of anti-realism that allows us to preserve the force of the social constructionist's position, while making clear how the accounts differ. I

will then make a case for the benefits of adopting the joint intentional mechanistic kind view over the conferralist account of the metaphysics of social categories by applying both to the case of sexual orientation. I have already argued that the account on offer more straightforwardly answers the appearance question than social positional views do (conferralism, while anti-realist, is still a social positional view of social kinds) while remaining neutral on the nature of social kinds. Therefore, I will focus on making the case that the joint intentional mechanistic view has epistemic benefits that outweigh those of conferralist account.

My aim in this chapter is threefold. First, I will argue against the claim that *metaphysical realism* is an important component of the debunking project by demonstrating the viability of two forms of critical anti-realism: the conferralist view (Ásta, 2018) and the joint intentional mechanistic kind view. Second, in doing so I will make clear how the joint intentional mechanistic kind view differs from the conferralist view: both are forms of critical anti-realism but the conferralist view is a social positional view and the joint intentional mechanistic kind view is not. Third, I will make a case for favoring the joint intentional mechanistic kind view by applying both accounts to the case of sexual orientation and arguing that the joint intentional mechanistic kind view better captures the case and the surrounding phenomena. I will conclude by discussing how the view on offer avoids the problems with social positional views outlined in chapter one and by responding to some possible worries about the joint intentional mechanistic kind view.

## THE VIABILITY OF CRITICAL ANTI-REALISM

Social constructionists tend to agree on a central point: race, gender, and other putatively social kinds are in fact real, but not biological kinds (Mills, 1998: Ch. 3; Mallon, 2004). Rather, they are often held to be social positions, and no less real for that (Haslanger, 2012: Ch. 7;

Barnes, 2017). But interpreting the different uses of “real” is notoriously fraught. Is the claim just that such social positions are as real as anything else? Scientists may talk about the differences between amphibians and reptiles. Does it follow from this that they believe there is a readymade division between the two, before each term was introduced as a useful shorthand for referring to relevantly similar animal species? If the realist and the antirealist can agree that our race and gender terms are useful and that we can make true generalizations about people belonging to race and gender categories, do they disagree on anything crucial? The crucial thing to agree on seems to be just that the regularities observed among members of such categories are not sufficiently explained by reference to the intrinsic properties of the relevant individuals alone, but thoroughgoing antirealists are perfectly capable of agreeing with this. Or so I shall argue.

Charles Mills understands realism, in the context of the sciences, as denoting the view “that entities postulated by natural science either exist or do not exist independently of human consensus or dissent” (Mills, 1998, p. 45) and as “associated with a belief in natural kinds with defining essences” (Mills, 1998, p. 46). But as we have seen, constructionists have declined to move from the realization that biological racial realism is false, to the conclusion that race is not real in other senses (Mills, 1998, p. 47). Thus, Mills suggests that we use the term *objectivism* to include the view that there are indeed races, but that they are determined by intersubjective agreement, and so have a sort of objectivity to them.<sup>66</sup> As we have seen, the social-positional theorist tends to go one step further and suggest instead that there is just a social position in a

---

<sup>66</sup> Mills writes “that the lines of demarcation, the categorical boundaries, are drawn here rather than there is a social decision, and one that creates the (social) reality in question” (Mills, 1998, p. 47-48).

social structure, a structure as real and—in a sense—perhaps almost as close to mind-independent as the atomic structure of table salt.<sup>67</sup>

Sally Haslanger (2012: Ch. 1) argues that the radical social constructionist makes a bad inference from the claim that some kinds are constructed to the claim that all kinds are constructed, and that doing so takes away from the force of the social constructionist's position. But what if one is simply already sympathetic to the broadly Kantian view that *all* descriptions of reality are in some way dependent on the ways we must represent the world given the sorts of minds we have? What if one were to hold that all kinds, even kinds considered paradigmatically natural kinds, are best understood as being determined by intersubjective agreement, or something else about us? Is it really true that this person cannot preserve the force of the social constructionist's position? Or is it in fact possible for one who holds this type of position to make sense of the special type of contingency at play in the cases of interest to the debunker?

In the last chapter, I argued that individuation conditions are not what pick out the kinds of primary interest to the debunking social constructionist. Here, I will emphasize that even if one were to hold that all kinds, even paradigmatically natural ones, were determined by intersubjective agreement, or some other thing about us, such as our cognitive apparatus, they could still make sense of the special type of contingency of interest to the debunker. The crucial thing to agree on is just that the regularities observed among members are not explained by shared intrinsic properties, but rather by shared intentionality directed toward a cluster of properties we happen to have an interest in. We can retain the distinction needed for social change regardless of our underlying ontological commitments. That is because there is a

---

<sup>67</sup> This comparison is a reference to Kate Ritchie's (2020) discussion of social structure.

different mechanism that explains the special type of contingency in the cases of interest to the debunker.

## THE CLAIM THAT CRITICAL REALISM IS *NECESSARY* FOR SOCIAL CHANGE

### *Background*

Social constructionists, and in particular those who claim that there are social kinds, are often associated with views that are typically thought of as radical or anti-scientific, such as idealism, relativism, or anti-realism about natural kinds. The social constructionists I have in mind are those who have been called “debunkers” (Haslanger, 2012, p. 92). Debunkers are not interested in kinds that are *obviously* social in one sense or another, rather, they are interested in kinds that *appear* to be natural but may in fact be wholly dependent on our social practices or, even more broadly, our conceptual activity in general. They may also be interested in kinds that are plausibly already considered social but may in fact be more social than has been widely or properly recognized (Haslanger, 2012: 132-136; Ásta, 2018: 36-37).

Artifacts and institutions are clearly causally dependent on us to be the way that they are, in that social factors play a causal role in bringing them into existence, and while this is an interesting topic in its own right, debunkers are more interested in uncovering kinds that may be unjustly considered natural, and thus inevitable, when they are not. As Haslanger writes in response to Ian Hacking’s (1999) idea of *causal construction*, developed in *The Social Construction of What?*, “we should avoid conflating social kinds with things that have social causes” (Haslanger, 2012, p. 131). The kinds of central interest are those that are suspected of being *constitutively dependent* on us. The suggestion is that properties such as “being a woman” or “being Black” are or might be meaningless in isolation. There are two senses in which such properties might be considered meaningless in isolation: 1) in that it seems that properties such



as “being Black” cannot be defined without reference to a position within a social matrix (Haslanger, 2012) and 2) in that it is the *perception* of Blackness, and not the actual intrinsic or natural properties of the person, that accounts for the constraints and enablements that are imposed on a person who is taken to have the property in a context (Ásta, 2013). The general idea that is being put forth in both formulations of the problem is that even when there is some supposedly natural or seemingly objective property in the vicinity, e.g., anatomical or phenotypic or other physical characteristics, these supposedly natural properties can come to have *social significance*, to be associated with other properties with which they are not clearly related, e.g., intelligence, trustworthiness, etc., and thus those who are taken to have these supposedly natural characteristics can come to constitute a social kind (which often parades as natural).

Sally Haslanger has proposed a middle path between two extremes—strict ontological naturalism on the one hand, and radical social constructionism on the other—by developing a view she terms *critical realism*. Before describing this view and her motivation for it, however, it will be helpful to go into more detail on what exactly is at stake in the debate on social construction, what positions are available, and what their respective strengths and drawbacks are according to Haslanger. The terrain is a treacherous one, because in the first place there appears to be widespread disagreement among philosophers over what it is to say, in a metaphysical debate, that something is “real,” and whether this implies that it “exists” and if so, in what sense.

#### *Strict Ontological Naturalism and the Error Theorist*

Many naturalist philosophers seem to think that to say that something is real in this context is always to say that it exists, where this is taken to imply an ontological commitment. Hence the tendency to reduce: Sure, x “exists,” in the loose sense, but it is *really* this other thing, e.g., an aggregate of microphysical particles. Consequently, for them it seems that the

significance of the metaphysical claim that something is socially constructed is always meant to be that it is not natural, or not objective, and thus that it is unreal. There are two positions available to this kind of philosopher. The tough-minded naturalist will take it that the metaphysical claim is simply confused; sure, our *idea* of the kind is socially constructed, but that does not mean the *kind itself* is (Ásta, 2013 characterizing opposition, p. 34). The more flexible and sympathetic of them will take an error theorist stance, they will not see the metaphysical claim as confused, and will follow it through to its apparent consequence: that because a kind is socially constructed, it has no basis in reality—it does not exist, and talk of it should be eliminated (see Haslanger, 2012, p. 299 for discussion of race eliminativism).

Although Haslanger does not engage too much with the first position, because it does not acknowledge the significance of the social constructionist claims whatsoever, it is important to note that it seems to be the default position, which all forms of social constructionism are a response to. The second position, that of the error theorist, she does engage with, because as she points out, it is much in the same spirit as the social constructionist position, although she argues that it is ultimately misguided. First, because it is hard to make sense of how to change the understanding of something that does not exist; second, because it is puzzling how it is possible that we seem to successfully refer to this thing on a regular basis. For those less concerned with maintaining a strict naturalistic ontology, where all top-level phenomena can be reduced to bottom level phenomena, the claim that something is real may not always be the claim that something exists, in the objective, naturalistic sense. It could be that top-level phenomena are constituted by, but irreducible to, bottom level phenomena, so that what is socially constructed is still real, still exists, and is still natural, *just not in the way that we thought it was*.

*The Focus on Constitutive Social Construction as a Response to Hacking*

Social Positional theorists suggest that *constitutive* social construction, a relation of metaphysical dependence, is the important notion/crucial relation for the purposes of the debunker. The suggestion that we focus on relations of metaphysical dependence arises as Sally Haslanger's response to what is found lacking in, e.g., Ian Hacking's account of the different types of social construction claims, all of which he interprets as *causal* claims. In his 1999 book, *The Social Construction of What?*, Hacking gives an analysis of what is meant when it is claimed that some x is socially constructed. Of particular interest is his discussion of *object-construction*, on which some objects, to our surprise, are the product of social forces: in particular, people of *certain kinds*. People can come to have, partly as a result of being categorized in a certain way, a set of features that qualify them as a member of a certain kind or sort (Haslanger, 2012, p. 122-123). On Hacking's account of what he terms *interactive kinds*, there is an interaction that happens through the *awareness* of the thing classified, (although mediated by the "institutions and practices surrounding the classification") (Hacking, 1999, p. 103, p. 31-2, 103-106; Haslanger, 2012, p. 123). Examples given of interactive kinds include women refugees, child viewers of television, child abusers, schizophrenics, as well as those labeled with various mental illnesses.

Hacking focuses much of his discussion on the way that *being categorized* in a certain way shapes the behavior of category members: it does so by shaping one's self-conception and thus one's behavior. Another focus of Hacking's discussion of social construction is *non-inevitability*. The contingent histories that led to certain idea-constructions are non-inevitable: the idea-constructions could have been different, and the corresponding self-conceptions could have differed accordingly, thereby shaping the behavior of those so categorized differently than they

in fact did. Sally Haslanger has responded that this notion of social construction does not capture what we need to capture with the notion of social construction. She argues that for the debunker, the claim of social construction is not simply that things could have gone differently, that, e.g., femininity and womanhood are understood in certain ways because of historical developments that could have gone differently. What needs to be captured, according to Haslanger and other social positional theorists, is that to be, e.g., a woman, *just is* to be constrained and enabled in particular ways: it is about what one can and cannot do or get away with—an insight attributed to Simone de Beauvoir (1949), when she wrote that “one is not born, but becomes a woman”—and the social positional theorist argues that one need not be *aware* of anything in particular in order to be so constrained or enabled (Haslanger, 2012).<sup>68</sup>

#### *Contrast between Thin and Thick Constructionism*

To understand the claim being made by the social positional theorist and its significance, it helps to again contrast it with another view in the vicinity: *thin constructionism*. The social constructionist sometimes adopts the view that the human kind in question, while not biologically real, is real in the sense of being, at the very least, *epistemically* objective (Mills, 1998: Ch. 3; Hacking, 1999: Ch. 1). For instance, one can understand *thin constructionism* as referring to the view that a kind is social when it is *ontologically* subjective but *epistemically* objective; when the kind in question is plausibly considered ontologically significant (i.e., in some sense *real*) but this significance (i.e., reality) is rooted in something like what Mills terms *intersubjective agreement* (Mills, 1998: 58). The thin constructionist about race or gender holds that there is an objective fact of the matter (of a sort) about what race or gender or other social

---

<sup>68</sup> Haslanger writes in responding to Hacking “But focusing on this process makes it seem that the impact of social forces on us and the locus of social change is primarily cognitive: social categories are offered to us that we internalize and modify, offering back a revised classification that others then adjust to (or not). Disrupt the classifications and you disrupt the social structure” (Haslanger, 2012, p. 127).

kind one belongs to, but that it is a sort of objectivity which is rooted in intersubjective agreement on Mills' view, or, as on Mallon's view, in common knowledge representations of the category (Mallon, 2016, p. 67). The social-positional theorist goes a step further than the thin constructionist, and instead endorses a *thick* constructionism about socially significant categories. The *thick* notion of construction favored by the social positional theorist is one that insists there is *no independent fact of the matter* (intersubjective or otherwise) about which kind one belongs to, *outside the way one is viewed and treated*. To belong to the kind is to be *taken to* have certain properties and be marked for certain kinds of treatment according to the background ideology (Haslanger, 2012, Ch. 7). The properties one *actually has* are deemed irrelevant because the kind of interest to the social positional theorist is the one composed of those constrained and enabled in particular ways due to perceived difference (Ásta, 2018, Introduction).

### *Critical Realism as a Response to the Radical Social Constructionist*

Social positional theorists suggest that it is important to social change that we be *metaphysical realists* both about perceptually individuated social positions *and* about non-social kinds. This insistence on what Haslanger terms *critical realism* arises partially in response to the views of the *radical social constructionist*. The *radical* social constructionist holds that *all* distinctions are made according to our interests, and so do not capture "nature's joints". The radical social constructionist seems to agree with the naturalistic realist that "existence," in the robust sense, depends on objectivity or mind-independence. However, they do not think that anything meets these criteria. All kinds are dependent on us, and thus objective reality is an illusion. Either everything belongs in the ontology, but it is a wholly *social* ontology, or nothing

belongs in “the ontology,” because independent reality is a “fiction”. Why take such an extreme stance?

For Catharine Mackinnon, on Haslanger’s interpretation of her, because the fact that some things and understandings are—upon further investigation—clearly dependent on us, implies that all other things and understandings that appear to be objective are not so objective, since it is the same mechanism in us that produces each of them. Haslanger attributes to MacKinnon, the view that “the power that has determined gender categories is the *same power* that has determined all categories” (Haslanger, 2012, p. 78) and a corresponding skepticism about sorting things according to their supposed natures or essences, perhaps even when those essences are qualified as *social* or *relational* essences, which one might view the social positional theorist as proposing we do. Once we have acknowledged that some kinds that appeared to be natural, and some understandings that appeared to be objective, are in fact constitutively dependent on our social practices, or guided by our interests as social beings—and if we also think that *all* of our concepts are a product of, as Haslanger puts it, the *same mechanism* or “pattern of thought,” which hopes to sort things and situations according to their objective natures, but may in fact always inevitably be guided by its own interests—are we not forced to question whether there are truly “natural” kinds at all? According to this line of thought, we might rightly say that “Reality” as a whole is *constructed* rather than *discovered* by our inquiry.

For Judith Butler, on Haslanger’s interpretation of her, it is because the fact that we inevitably make *distinctions* with our own interests in mind, and so for social purposes, implies that the *differences* in question are only social differences, that is, they have no objective basis. For example, the radical social constructionist is likely to reject the claim that “gender is the

social meaning of sex,” because they hold that *sex itself* is constructed according to our interests. It is not that sex is one of nature’s givens, and gender is the social meaning of having a certain biologically determined sex. According to the radical social constructionist, we fashion the sex binary to meet certain ends of ours. If sex is already constructed, it is of little use to even distinguish between sex and gender (Butler, 1990: 7; Hacking, 1999: 8-9).

Haslanger argues that both commonly made inferences in the social constructionist debate are fallacious. She warns against jumping from the idea that our *point of view* on the world is socially mediated to the conclusion that the *world itself* is socially mediated, or from the idea that we make *distinctions* for social purposes to the conclusion that they correlate with no objective *difference* in the world. She argues that because these inferences are faulty, they need not commit us to idealism, relativism, or anti-realism, and argues instead for what she calls *critical realism*. In fact, she has argued that social constructionism not only *need* not commit us to such radical, anti-scientific views, but that it *should* not. Why? Because on her view, when philosophers such as Mackinnon and Butler make the inference—which she argues is a faulty one—from the claim that *some* kinds are dependent on us to the claim that *all* kinds are, or from the claim that we make *distinctions* for social purposes to the claim that the *differences* in question are always also social, they seem to implicitly make the claim that both the robustly social and the seemingly natural are produced by the *very same mechanism*, thus implying a new sort of determinism and inevitability. The critical realist’s demand for social justice cannot afford to allow such implications, just as it could not afford to accept the tough minded naturalist or uncritical realist’s claim that all kinds are independent of us, and thus determined and inevitable.

The social positional theorist claims something more radical than both the thin constructionist and the dynamic nominalist but does not want to go as far as the radical social

constructionist, or to be associated with antirealist or idealist views. The proponent of the standard social positional view will tend to insist that kinds unified by a social property are nevertheless real kinds, that social kinds are no less *real* for being social. There are also kinds that are determined completely independently, and retaining this contrast, the social positional theorist holds, is essential to social change. There is a spectrum, within the real, within the natural, from non-social to social (Haslanger, 2012: 213). We are capturing nature's joints when referring to social kinds, but they are joints we *made* (Barnes, 2017).

#### SKETCH OF CRITICAL ANTI-REALISM AND COMPARISON WITH CONFERRALISM

Haslanger's main goal, in the articles that I have been referring to is to show that social constructionism need not be associated with the views of the radical social constructionist. I agree with her that it *need* not be, and further I admire that she has cleared a path for naturalistic social constructionism, or as she has called it, critical realism. My only objection is to her further claim that anti-realism *necessarily weakens* the theoretical and political force of the social constructionist position. My objection arises from two related sources: 1) I agree with Haslanger that the inference from the idea that *some* kinds are dependent on us to the idea that *all* kinds must be dependent on us, *in the same way*, is fallacious. However, I think that if one finds plausible reasons elsewhere for being skeptical of realism about even natural kinds, one can still make a *distinction* that is necessary for inspiring social change: the distinction between natural and social kinds

#### *Essentiality Conferred*

Inspiration for the view I want to sketch can be found in Ásta's (2008) anti-realist essentialism, which voices concern over what the semantic externalist's thought experiments can really tell us about the nature of a thing. She defends a conferralism about essences, which can



be seen as giving a more sober and systematic account of Judith Butler's inference—the one critiqued by Haslanger—from the idea that we make *distinctions* for social purposes, or otherwise according to our interests, to the idea that the *differences* in question are social or are otherwise dependent on us (Haslanger, 2012; Butler, 1990). Ásta writes, “if a property chiefly figures in explanation of social facts, and not natural facts, that suggests the property in question is a social property, and not a natural property” (Ásta, 2011, p. 16). For example, if sex is taken to be grounded in natural facts about anatomical difference, but these “natural facts” factor primarily in explanations of rights and statuses, we have to wonder whether it captures a genuine “joint of nature” or rather, the only relevant joint. The proper classification, if we wanted sex to explain something else, such as reproductive capacities, would radically alter our idea of sex, since our current conceptions do not line up with these capacities (Ásta, 2013).

The main idea is that because it is always our interests that guide our classifications, we can never be sure about the true “nature” of a thing, or put another way, whether we are capturing nature's “joints”. Because our thought experiments can at best tell us how we use our concepts, they cannot show that we have captured a real essence of the kind in question. Clarity of our intuitions about what is essential to a thing only amounts to clarity of our conceptual commitments and does not in fact prove that the thing has a certain property essentially, independently of our interest in it. The critical anti-realist position that I am sketching would admit, as Kant did, that there is an underlying something, that there are constraints on the part of nature, but concede that the way it appears to us may not be the only way it could be represented. In some sense, we are always capturing nature's joints, but the ones we highlight are arbitrary, or rather, the ones that happen to matter to *us*.

My view is that one can hold this critical anti-realist position and retain the distinction between natural kinds and social ones, by pointing out that even if *all* kinds are cognitively or even socially *mediated* some are only genuinely explanatory, i.e., are only genuinely *kinds*, due in part to joint intentional mechanisms. One can acknowledge that all kinds are dependent on us to be the way that they are, in that all we can ever get at is how things appear to us, without implying that independent reality is a “fiction,” or that all kinds are dependent on us *in the same way*. All kinds are dependent on us to be the way that they are, but there are different ways of being so dependent, e.g., merely mediated, causally constructed, constitutively constructed, or, importantly, they can be explanatory as a result of joint intentional mechanisms. And none of this need imply that independent reality is a “fiction”. This view takes seriously Kant’s idea that while appearances do imply an underlying something, we cannot know what that something is like *essentially* independently of our interest in it.

Anti-realism about natural kinds does not necessarily distract from, distort, or weaken the theoretical and political force of the social constructionist position, because we can retain the distinction that is necessary for social change regardless of our underlying ontological commitments. While it is helpful to introduce the problem as a distinction between the natural and the social, it is important to eventually see that the important distinction is between the inevitable and the non-inevitable, and that therefore it does not actually matter what we think counts as “real” or what we think it means to “exist”. We can retain the inevitable and non-inevitable distinction whether our ontology is wholly social, wholly real, or some combination of the two. What is needed is to make as clear as possible that the regularities observed in the cases of interest are in a special sense contingent.

Commitment to real social positions seems to force one into a thoroughgoing realism. And as I have argued, they are not required to make the kinds in the vicinity explanatory, i.e., kinds. Instead, what is crucial is to show *why* they are explanatory, to reveal how they are social even though they appear natural. We can (and I think we should) capture the core phenomenon (when a kind appears natural but is not) without bifurcating ontology in the way that is often suggested by the social positional theorist. Because what is needed is just to locate a different mechanism. On my view, it could be that all kinds are perceptually individuated (or, at least, in some sense “dependent on us”), or it could be that none are. That is because, on my view, individuation conditions are not what distinguishes social kinds from natural ones. Thus, a commitment to social constructionism about certain socially significant kinds does not require that we accept a thoroughgoing (although critical) realism.

#### *Conferralism vs. The Joint Intentional Mechanistic Kind View*

There are, however, important differences between the conferralist view and the joint intentional mechanistic kind view. When Ásta turns specifically to a conferralist view of the metaphysics of social categories, she has sometimes started with the particularly vivid example of baseball facts first. Before we apply both accounts to a more pertinent example, we can begin to see the distinctions with the baseball case. Take Ásta’s example of a pitch being a strike. On the conferralist account, the umpire is attempting to track the trajectory of the ball (this is the grounding property) but it is the *umpire’s judgment* of that trajectory that determines whether the pitch is a strike: the actual trajectory of the ball does not matter to being a strike, ultimately. Being a strike is a conferred property, i.e., a social property. This can serve as an analogy for how a social status is conferred. On the conferralist account of the metaphysics of social categories, being taken to have some grounding property P in a context results in the conferral of

a social status that consists in certain constraints and enablements. This is what it is to belong to a social kind. For example, being judged to have a certain sex assignment, the relevant grounding property in a certain context, may result in the conferral of the status *woman*, and with this, certain constraints and enablements such as access to free drinks at the bar.

On my view, it is not the conferral, or “taking to be” that makes a phenomenon in question social. If it is social, it is because the causal role that something plays (e.g., the pitch being a strike) is *actually* played by the joint acceptance of the belief that this particular combination of phenomena amounts to whatever the implications of a strike are agreed to be in that situation. There is the trajectory of the ball, and the umpire’s vision and judgment, but these are just phenomena in the world. Our attitudes about this combination of events are what explains the consequences of the pitcher moving his body in a certain way and the umpire moving his mouth in a certain way. Being a strike is a social property because the relevant phenomena are only explanatory of certain outcomes due to our joint commitments about what ought to happen whenever that combination of phenomena is observed. Similarly, we can explain the outcome that those who meet the descriptive component of the jointly accepted representation of the kind woman get drinks at the bar insofar as we are jointly committed to representing the kind women in a covertly normative way, such that one might be rebuked for not providing them.

#### APPLICATION TO A CASE: SEXUAL ORIENTATION

Analytic metaphysicians have recently turned to analyzing what it is to have a sexual orientation. Interestingly, very few seem tempted to talk about sexual orientation as a social position. This seems worthy of investigation given the ways that sexual orientation seems analogous to race and gender in certain important respects. Ásta has suggested that this is

because the projects I have in mind can be understood as analyzing or giving an account of the *grounding property* we are attempting to track when conferring a social status, e.g., the grounding property might be *identifies as a woman and desires sex with women*, and the social status might be *lesbian*. In this section I make a case for thinking that looking at the case of sexual orientation shows the strengths of the joint intentional mechanistic kind view and that the joint intentional mechanistic kind view better accounts for the social construction of sexual orientation than the conferralist account.

### *The Main Accounts of Sexual Orientation*

Until recently, very few analytic philosophers had considered questions about the nature of sexual orientation: what sexual orientations are, or what ‘sexual orientation’ means (Diaz-Leon, 2022).<sup>69</sup> Robin Dembroff’s (2016) paper was the start of a growing debate within analytic philosophy on what sexual orientations are, or as it is sometimes put, which concept of sexual orientation we are *actually* using and which concept we *ought* to use (Diaz-Leon, 2022). One common sense approach to understanding sexual orientation is to understand it solely in terms of behavior. A behaviorist about sexual orientation holds that a person’s sexual orientation is determined solely by their observable sexual behavior. However, as noted by Diaz-Leon (2022), this sort of view fails to acknowledge that some individuals, such as voluntary celibates, might never act on their sexual orientation. It also fails to acknowledge that some individuals may engage in certain sexual activities due to societal pressures rather than as an expression of their sexual orientation (Diaz-Leon, 2022). Thus, the starting point for the current debate might be Stein’s (1999) *dispositionalist* account, on which a person’s sexual orientation is determined solely by what sex(es) and gender(s) of persons they are *disposed* to engage with sexually (under

---

<sup>69</sup> Diaz-Leon (2022) cites Stein (1999), Calhoun (2002), Wilkerson (2013), Dembroff (2016) as exceptions.

certain manifesting conditions). The ideal conditions are those where no forces prevent or discourage a person from acting on their desires (Stein, 1999, p. 45).

Robin Dembroff (2016) puts forth more specific manifesting conditions in their argument for *bidimensional dispositionalism*, on which a person's sexual orientation is grounded in their dispositions to engage in sexual behaviors under ordinary conditions for these dispositions. More specifically, which sexual orientation a person has is grounded in what sex(es) and gender(s) of persons they are disposed to engage with sexually under those conditions (Dembroff, 2016, p. 18). The ordinary conditions proposed by Dembroff are those in which attraction to persons of a certain sex or gender is due (at least in part) to their being that sex and/or gender, in which there is a reasonable diversity of potential sexual partners, and in which one is willing and able to sexually engage with other persons (Dembroff, 2016, p. 17).

Diaz-Leon raises some objections to Dembroff's view, and to dispositionalism in general, by offering counterexamples where such an account seems to get the wrong result. For example, she considers the case of a woman who identifies as bisexual and who has decided to be in a long-term monogamous relationship with her male partner. Diaz-Leon argues that Dembroff's bidimensional dispositionalism has the result that this person is heterosexual, when the result should be that she is bisexual. She argues that for the familiar reasons behaviorism was doomed to fail, sexual desires cannot be reduced to the behaviors a person is disposed to engage in under ideal conditions, independently of the mental states of the person. She proposes that we understand sexual orientation in terms of sexual preferences understood as complex mental states, i.e., dispositions to instantiate certain desires and feelings. She calls her view desire dispositionalism, the view that one's sexual orientation is determined by the sex(es) and

gender(s) of persons for whom one is disposed to have sexual desires under the relevant manifesting conditions (and one's own sex and gender) (Diaz-Leon, 2022).

Finally, Matthew Andler (2019) argues that a stalemate occurs between these two views because both assume the priority of intuitions about the concept of sexual orientation over intuitions about the taxonomy of sexual orientation. He argues that intuitions about which groups fall under the concept sexual orientation should be prior instead. More recently, Andler has argued further that mere sexual dispositions do not count as sexual orientations unless/until they are “indexed to social practices that privilege, subordinate, and/or marginalize individuals in relation to heteropatriarchal kinship structures” (Andler, Forthcoming). This is the sort of view I was looking for, given the ways that sexual orientation seems analogous to cases of race and gender. If social constructionists want to define race and gender in terms of social position, why not do the same with sexual orientation? As mentioned, one way of answering this question is to understand the first two proposals, at least, as providing an analysis of the grounding property we are attempting to track in the conferral of a social status.

#### *How the Conferralist Handles the Case*

On Ásta's account, the accounts of sexual orientation just discussed would actually be accounts of the grounding property for the conferral of a social status. Her recommendation is that all these views, or at least the first two, can be understood as determining the grounding property for the conferral of a social status. The trait in question would be the grounding property being tracked, and the individuals that have a social status, consisting in certain constraints and enablements, conferred on them as a result of being taken to have the grounding property belong to the relevant social kind. The grounding property could be many different things depending on the context. The grounding property in a context might be a role, a desire

for certain sorts of sexual activity, a disposition, or an identity. The social kind would be unified by the social status conferred onto those taken to have the relevant grounding property in a context. So, *lesbians* understood as a social kind might be those who share a social status in a context due to being taken to have a desire for certain sorts of sexual activity. To be a lesbian, then, might be to occupy a certain position in a context as a result of being taken to have a desire as a woman for sex with other women.

*Various Objections to Social Positional Accounts in General and the Conferralist Account in Particular*

Before going into how the joint intentional mechanistic kind view handles the case, I will first discuss some worries about the results of the conferralist approach and the results of social positional approaches in general. Some worries about social positional approaches in general include inclusion and passing, intersectionality, and positive identification. Some worries about the conferralist approach in particular are proliferation, and systematicity.<sup>70</sup>

The conferralist, like the social positional theorist in general, suggests that the members of the social kind are just whoever is subordinated or privileged based on presumed characteristics. Notice that just as in the case of gender, where the result is that membership in the kind *woman* depends on being regularly or for the most part *taken* to have female sex characteristics, it seems that not all those who identify as women will count as women (Jenkins, 2016), in the case of sexual orientation those who identify as having a certain sexual orientation can fail to count as belonging to that sexual orientation. While focusing on the case of gender,

---

<sup>70</sup> I do not discuss the notion of intersectionality in detail but only raise worries about the ability of social positional accounts to account for it. I hope to do more justice to the topic in the future. See Kimberle Crenshaw (1989, 1991) who argues that an additive model of identity is inadequate and proposes an intersectional analysis instead. Also, some of the worries I raise in this section are addressed by for instance Ásta, 2018 and Haslanger, 2012. I raise them still because I suppose I am not rid of the worries, and I wish to address them in more detail in future work.



Katharine Jenkins (2016) has referred to this as *the inclusion problem*. It is important to note that for the conferralist, the status one has, and the grounding properties being tracked in the conferral of that status, will depend on the context. On either approach, however, the result seems to be that one can fail to belong to the sexual orientation one identifies as having.

Relatedly, the notion of passing seems not easily made sense of on this kind of account. If belonging to the social kind *lesbian* is to have a social status in a context (or to occupy a position in a hierarchical social structure, as on the Haslangerian sort of view) due to being taken to have the relevant grounding property, then if one is not taken to have the grounding property, one does not have the status and does not count as a member of the social kind. There is no such thing as being taken to be straight and yet *not* being straight, because being taken to be straight gets you the social status of being straight.

The social positional theorist will likely respond that belonging to the social kind *lesbian*, or the social kind *woman*, is a different matter than having the sexual identity *lesbian* or the gender identity *woman*. Haslanger and other social positional theorists are interested in a concept of gender that has members of a social class as its extension. If a person does not pass as a woman, and so is treated as a man, then in theory, they are unlikely to be in a subordinate position. In response, Jenkins emphasizes the harm done to trans women by a feminist analysis of gender concepts on which trans people do not belong to the gender categories they identify with (Jenkins, 2016, p. 396). And she proposes that we retain Haslanger's concept of gender as an imposed social class but also that two target concepts of gender are required for our theorizing. The other concept "captures the sense of gender as a lived identity and draws on Haslanger's (separate) account of racial identity" (Jenkins, 2016, p. 397). As the quote suggests,

Haslanger has entertained that there could be positive racial and gender identities.<sup>71</sup> However, she does maintain the position that the topic of race and gender identities is a separate topic from what race and gender themselves are (i.e., positions in a social structure). My thought is that this is even less convincing in the case of sexual orientation than it is in the case of gender or race.

Another question is whether it is possible to positively identify with a kind that is analytically oppressed. It seems that there will be many who want to be, e.g., women, lesbians, and so on “after the revolution”. Furthermore, is there not something that, e.g., femmes, bisexuals, queers, autistic people, and so on *really share*? Not an essence, on my view, but something on the basis of which they prefer to categorize themselves in these ways? It may be useful to talk about the way that people who belong to these categories are positioned similarly, but I think it is possible that making what are *to the “folk”* identities into *essentially* social positions, whether conferred or not, whether in a context or in a social structure, may cause more harm than good, or at the very least resist uptake.<sup>72</sup>

There is also a concern about whether positing social positions or statuses is consistent with the insights of intersectional feminism. A theory that posits stable positions that are as coarsely grained as *woman* seems lacking in empirical adequacy. Intersecting identities dictate and give rise to differences in privilege and disadvantage among members of these broader categories. It becomes unclear what grounds we have for positing social positions and at what resolution we ought to do so, if at all. Returning to conferralism and the case of sexual orientation, if we think about conferring a social status in a context, will all of those taken to have the grounding property of, say, a disposition to engage sexually with women all be given

---

<sup>71</sup> See also Alcoff (2006) for a discussion of racial identity.

<sup>72</sup> It may be responded that acknowledging the similarly oppressed individuals as a kind does not prevent the existence of or development of a group that positively identifies as having something in common, see Ritchie (2021)

the same status, *lesbian*? Is it not more likely that there will be different constraints and enablements conferred even among those taken to have this grounding property? Different constraints and enablements for those who are taken to also have the grounding property of, for instance, having a certain physical disability, or ancestral links to a certain geographical region, or relationship style, and so on, in addition to having the disposition being tracked?<sup>73</sup>

It seems there is a need for extremely finely grained clusters and corresponding nodes, and perhaps some social positional theories, such as Ritchie's (2020) framework can provide this. But even then, the idea that two individuals could share the exact same social position seems dubious. A theory that posits stable and coherent social positions that are as coarsely gained as *woman* seems lacking in empirical adequacy. As we saw above, intersecting identities dictate and give rise to differences in privilege and disadvantage among members of these broader categories. It becomes unclear what grounds we have for positing social positions and at what resolution we ought to do so, if at all.<sup>74</sup> We might ask: in what sense does every person who is taken to have a certain role in biological reproduction, say, occupy *the very same* social position? Once again, Haslanger and other debunkers are not unaware of this issue. Haslanger builds into her definitions of race and gender that members are privileged or subordinated *along*

---

<sup>73</sup> One worry I would like to address in future work is that we seem to have representations of single moniker kinds. If this is the case, it might be a problem to recommend that we simply add up the different single moniker kind representations, if one wants a restraint on the account that it can do justice to intersectionality. However, it seems to require empirical investigation whether it is true that we only have representations for single moniker kinds. But if this is the case, one route is to say that sometimes there is a single moniker kind that actually could be understood as the combination of two or more categories. Take, e.g., the social category *femme*. A femme, I take it, is a queer person who identifies with stereotypically feminine gender presentation. Thanks to Kate Ritchie for pressing on this point.

<sup>74</sup> One response to consider is why coarse-grained classifications are warranted in other cases. For example, there are lots of differences between animals. Should we only posit very specific kinds, e.g., *siberian tiger*, since it is clear that Siberian tigers and American crocodiles are very different? Is positing broad kinds like *animals* or *mammals* unwarranted? I am inclined to think that because the kind is said to share a social position, there should be a single social position, but more consideration of the difference between these cases is needed. Thanks to Kate Ritchie for this question.

*some dimension* on the basis of presumed characteristics. Furthermore, Haslanger acknowledges that belonging to a given social kind can privilege or subordinate depending on context. She writes:

“for example, in the contemporary US, there are contexts in which being Black and male marks one as a target for certain forms of systematic violence (e.g., by the police). In those contexts...being male is not something that a man ‘has going *for* him’; though there are contexts (also in the contemporary US) in which Black males benefit from being male” (Haslanger, 2012, p. 232).

Being a man seems to add to rather than alleviate the disadvantage of being Black when it comes to police brutality. However, a Black man might be more likely to be taken at his word than a Black woman. The point is that it seems undeniable that all humans will fall into more than one social kind, and the intersecting disadvantages and advantages seem to require their very own social position if the view is to be coherent. It seems that there must be countless more finely grained positions within any more coarsely grained position that might be posited. This makes implausible the suggestion that there are real, stable, perceptually individuated social positions shared by all who are taken to have some feature. It seems inaccurate, and perhaps unethical, to suggest that a wealthy, white, able bodied, cis gendered, heterosexual woman and a poor, disabled, trans woman of color share a social position.

Perhaps the conferralist could respond that in some cases the grounding property for the conferral of a social status is a cluster of properties that is intersectional, e.g., maybe in a context there could be a social status *Black femme*, and the grounding property is in fact a cluster of properties including some combination of gender identity, gender presentation, racial identity and ancestral links to a certain geographical location, say. This seems plausible enough, however, there is a question of whether it is desirable that there is both a cluster of properties that is being tracked and a social status that is conferred, and whether this is necessary to capture

what debunkers want to capture. Furthermore, there is a question of a deeper sort of proliferation that can occur, one, because of the different statuses available in different contexts, and two, because of the possibility that the grounding property could itself be a social status, to which a social status is conferred. Is there going to need to be a social status that is conferred on those that are taken to have a certain social status in a context? Would it not be preferable if the possibility of this sort of proliferation could be avoided?

As far as objections to the conferralist approach as opposed to those directed at social positional views in general, I think there is a case to be made that out of the two critical anti-realist positions, conferralism and the joint intentional mechanistic kind view, the latter has an advantage in capturing the systematicity of category-based oppression. While the context sensitivity of the conferralist account has some advantages over the social positional view of the critical realist in capturing the varied ways we can be constrained and enabled from context to context, it might be that it has an even harder time than does the critical realist proposal in answering the appearance question. Because of the scope of the social structural account provided by Haslanger, an answer to the appearance question can perhaps be inferred although it is not straightforwardly given. Because the constraints and enablements are so varied depending on context in the conferralist account, the answer of why such kinds appear natural becomes even less straightforward. The joint intentional mechanistic kind view is able to resolve this question by providing the mechanism responsible for the regularities that we observe and which lead to the true generalizations from which we incorrectly infer a shared essence or nature.

#### *The Result on the Joint Intentional Mechanistic Kind View*

On my view, collections that share any of the described traits could be considered a social kind insofar as that collection features in regularities explained by joint intentional mechanisms

with a certain sort of content. My view is that sexual identity, or intersection of sexual identities which may come in degrees and on different axes, should be a primary consideration in determining sexual orientation.<sup>75</sup> I think one's sexual orientation is best understood as a coordinate involving a complex interaction between gender, social role, identity, disposition, desire, and relationship style. Any one of these facets of sexual orientation, or any complex coordinate of interaction between them, could individuate a social kind insofar as there are joint commitments to representing that cluster of properties in a covertly normative way. If one is a polyamorous bisexual woman, it seems incorrect to say that this is due to something about other people. However, it is still certainly true that some generalizations about polyamorous bisexual women will be true due to something about other people, namely, their joint commitments.

Whereas on the conferralist account there is a context in which being taken to have some grounding property results in the conferral of a certain status that is comprised of particular constraints and enablements, on the joint intentional mechanistic kind view there is a cluster of properties that a collection of individuals is jointly committed to representing in a way that is covertly normative. On the conferralist account, to say that sexual orientation is socially constructed is to say that belonging to the kind is a matter of being taken to have certain traits and having a certain status consisting in constraints and enablements conferred upon one as a result. On the joint intentional mechanistic kind view, to say that sexual orientation is socially constructed is to say that joint commitments with certain sorts of covertly normative content are the mechanism responsible for some of the regularities observed among members of the kind.

---

<sup>75</sup> See Richardson (forthcoming) for a scalar theory of social categories in which social identities come in degrees.

The tricky part is that in many of the relevant cases, notably cases involving traits that are widely held to be deviations, as in the case of certain disabilities, sexual desires, or relationship styles, the regularities observed among members of the kinds are often explained by the fact that individuals with such traits are excluded from a broader kind, e.g., *person*. One might think this is going on in most cases of interest to the debunking social constructionist, that full personhood is denied in the cases of, e.g., certain sex assignments or ancestral links to certain geographical regions and so on. However, at least in the cases of sex and gender, there are prescriptive norms attached to the representations of the kinds as well, perhaps in addition to failing to meet the requirements for full personhood.

When talking about being denied full personhood, I am returning to the issue discussed in response to the “too cognitive” objection in chapter two. The objection involves examples of cases where it seems that *features of the environment*—and so perhaps material practices rather than joint commitments—explain why some people get access to certain resources and others do not. For example, when the bus line being down disproportionately impacts individuals with lower incomes. I responded that joint commitments may play a role in shaping the relevant practices. However, those sorts of examples lead me to think that in some cases the relevant content or representation is not of the kind in question *being a particular way* but rather the members of the kind being *left out* of a representation of a broader kind that is understood as properly being another way. They bring to our attention the way that sometimes the content of a shared intention that explains regularities among members of a kind is “negative” in the sense that the regularities are a result of the members of the kind being excluded from a broader kind, such as *person*, that has a normative dimension not met by members of the particular kind of interest which is a subset of the broader kind. For example, the normative dimension of our

representation of the kind person might include being able bodied, and this would result in those who are disabled being denied full personhood.

Let us take some more examples. If we want to explain disproportionate levels of homelessness among queer youth in a society, it might be that this society represents the kind *person* in a covertly normative way such that while queer individuals meet the descriptive component of the representation of the kind person, they fail to meet the normative component of the representation that prescribes that persons should have the potential to desire heterosexual sexual experiences. This is an example of a “negative” case, where there is a failure to fit into a broader category. If we want to explain the disproportionate levels of domestic violence toward bisexual women in a society, it might be that this society represents the kind woman in a covertly normative way such that while bisexual women may meet the descriptive component of the representation, they fail to meet the normative component of the representation that prescribes that women should serve the sexual needs of men. This is still negative in the sense of failing to meet the normative component of a representation, but it is negative in a different sense since it is a failure to meet the norms for the more specific kind to which they meet the descriptive component.

## HOW THE PROBLEMS OF CHAPTER ONE ARE AVOIDED

### *Thoroughgoing Realism*

In chapter one, I argued that on the prominent approach to debunking, social kinds might be interpreted as kinds that share a social essence: an essence composed of a social property, that all and only members of that kind share. However, because not even natural kinds are thought by philosophers of science to share an essence, but rather to share a cluster of properties about which we can make reliable predictions, positing real social positions or essences did not seem to



rule out pernicious generalizations about kinds that share a cluster of properties that nevertheless are not explanatory in virtue of that cluster of properties. I argued further that insisting on the reality of social positions thus needlessly forced thoroughgoing realism, as it would be odd to be realists about social positions but anti-realists about tables and chairs. The joint intentional mechanistic kind view does not needlessly alienate anti-realists or require a thoroughgoing realism because the focus is on what is explanatory in the cases of interest, rather than on the nature of the kind. By focusing on what is in fact explanatory in the cases of interest, by giving a straightforward answer to the appearance question, it also does a better job of blocking pernicious generalizations.

### *Explanatory Target*

In chapter one, I argued that because the prominent approach to debunking is to argue that the relevant kinds are constitutively socially constructed, there is a failure to pick out the kinds of actual interest to debunkers, namely, covertly socially constructed kinds. Focusing on kinds that cannot be defined without reference to social factors has the result that the kinds of interest are lumped together with kinds like *money*, *president*, *judge*, *point guard*, and so on. The joint intentional mechanistic kind view recommends instead focusing on the explanation of the regularities observed among members of the kinds of interest, rather than on the nature of the kinds. I argued that joint commitments with certain sorts of contents are the mechanism that explains regularities among members of the kinds of interest and that figuring out what the relevant contents are allows us to single out the kinds of interest and explain rather than merely label the distinction between overt and covert construction.

## *Ontological Collapse*

Finally, in chapter one I argued that non-mentalistic accounts of the phenomena of interest, e.g., those that focus on countenancing real social structures, allow the phenomena of interest to collapse into natural structures. I made the case that on accounts that focus on constitutive social construction, such as Ritchie's (2020) structuralist ontology of social groups, the social structures within which social kinds are thought to occupy nodes are not sufficiently distinguished from perfectly natural structures. The mechanistic explanation I have provided aims to illustrate that the phenomena of interest does not straightforwardly collapse into more "fundamental" structures. I have argued that providing a mechanistic explanation allows for and perhaps encourages agnosticism about the relationship between levels of explanation, whereas a focus on relations of metaphysical dependence seems to result in a sort of epiphenomenalism about the phenomena of interest. But I have also emphasized that because plural subjects are material objects that have intentionality, they do not straightforwardly collapse into natural structures. That is because it need not be the case that all the individuals that are jointly committed are also personally committed.

As Gilbert (2009) has argued, "as a conceptual matter, when two or more people share an intention, none of them need have a personal contributory intention" (Gilbert, 171). Gilbert discusses a case where Ned and Olive form an agreement to hike to the top of a hill. Part of the way up the hill, Ned realizes it will be too much for him to try to make it all the way to the top. When Pam walks by and asks them how far they intend to go, Olive replies that they intend to go to the top. Intuitively the pair still shares this intention, even though Ned's personal intention does not reflect this. We can also think of cases where a group believes something that no single member privately accepts. For instance, if the department votes for a candidate and no one's first

choice is elected, then the department believes the winning candidate is the best pick for the job even if no individual member of the department thinks so (Gilbert, 1987; Tuomela, 2007; Pettigrove, 2016). In the same way, it is possible that members of a community could have an implicit joint commitment that is not a personal commitment of all members. For example, one's family may be jointly committed to the belief that the women of a family are to make dinner and clean up afterward. However, it is possible that individual members of the family are not personally committed to this belief, even while acting in accordance with it. If maybe even most of the members of the family do not personally accept this, it will be difficult to explain why the women of the family do indeed make and clean up dinner every evening, without reference to the joint commitment.

## RESPONSES TO WORRIES

### *The Extensions of Social Group and Social Kind*

There is sometimes a confusion about the extensions of social *kinds* versus social *groups* on my account. On my view, these are related but importantly distinct phenomena. I propose that a group is a *social group* if its members jointly accept a covertly normative representation of individuals with a certain cluster of features, and this acceptance explains some of the regularities observed among individuals with that cluster of features. I propose that a kind is a *social kind* if its members share a certain cluster of features, and if (at least some of) the regularities observed among members are explained (in part) by the joint acceptance of a covertly normative representation of the kind associated with that cluster of features. The members of the social kind may very well be members of the very social group or plural subject that plays a role in explaining the extent to which that social kind features in regularities. For instance, a woman may very well be party to the joint acceptance of a covertly normative

representation of the kind *woman*, and thus be a part of the plural subject that partially explains regularities that obtain among women. On my view, women do not form a social group. They form a social kind. They may come to form a social group by jointly accepting a rule, belief, norm, etc. But it is unlikely that such a group would be co-extensive with a social kind.

Say we want to pick out with the term *women*, those who sincerely self-identify as women. Is it conceivable that each one of the individuals picked out could be parties to a joint commitment? Recall that I am using *social group* and *plural subject* synonymously. I do not oppose the use of other senses of *social group*. Social groups understood as plural subjects, Gilbert writes,

“can be large, long-lasting, hierarchical, impersonal, and anonymous: the individual members need not know each other personally, or even know of each other as individuals. Certainly, they need not have sat down round a table and hammered out an agreement for all to sign” (Gilbert, 2018, p. 226).

However, they need have expressed willingness (either implicitly or explicitly) to act as a body that, e.g., believes that P (Gilbert, 2013, Introduction). It is doubtful that all and only those who sincerely self-identify as a woman will ever be party to a single joint commitment. It is thus doubtful that they could form a social group in my sense. However, it seems to me conceivable that all the members of this same category could feature in regularities that are explained in part by the actions of a plural subject, i.e., by joint intentional mechanisms, and thus that they constitute a social kind. Plural subjects, on my view, can be thought of as the subjects of ideology and of conceptual frameworks. I’ve argued that it is useful to think of these as paradigmatic social groups and to contrast these with social kinds (cf Gilbert, 1989: Ch. 4). But my goal is not to prevent people from using *social group* (or *social kind*, for that matter) in other ways. The important thing is that I am giving an analysis of the subjects of conceptual schema

and ideology and showing how these relate to what are on my view paradigmatically social kinds.

### *The Role of Representation*

One may ask whether, on my view, social kinds are kinds before they are represented. Because I am focused on kinds that seem to be natural but are social, I am focused only on kinds that we have representations of (since we have views about them being natural). Thus, I am able to avoid the worry that there could be social kinds that are not represented. On my view, there is likely to be an identifiable cluster of properties shared by individuals belonging to the kind in question. I would avoid giving an answer to the question of whether it is a kind at all before it is represented: it *may* be a kind if there are regularities among individuals with the cluster of properties before it is represented. But it is not a *social* kind until there are joint commitments pertaining to the individuals with the relevant cluster of properties, and which partially explain regularities in which the kind is featured.

### *Commitment to a Representation*

One may very well find odd the suggestion that one could be *committed* to a representation let alone *jointly* so committed. On the Gilbertian account of what it is to share an intention, it is to be jointly committed to acting in accordance with a given intention, belief, assumption, rule, and so on. One way that we know there is a joint commitment to acting in accordance with some attitude is when we are aware (to some extent) that *not* acting in accordance with it could result in a *rebuke*. Admittedly, a commitment to a representation is an odd thing to consider. But we may think of common rebukes we may hear in everyday life, e.g., “ladies cross their legs,” and “boys don’t cry”. These are presented as factual descriptions or generalizations made by uninterested parties. Similar to “penguins don’t fly” or “mosquitos drink

blood”, we are invited to think that the properties shared by the individuals to which the term applies *themselves* explain the regularity noted. When in fact, in the former case it seems self-evident (again, upon reflection) that the prevalence of the statement itself contributes to the truth of the generalization. Individuals in a given interaction may personally reject that ladies or boys share essences or are natural kinds, and still feel compelled to either rebuke those who do not act in accordance with the representation or else to behave in ways that comply so as to avoid rebuke themselves.

My view is not that being rebuked for crying as a boy or failing to cross one’s legs as a girl will somehow determine one’s gender, by forcing one to act a certain way or insisting that one has failed to meet certain expectations. Rather, it is that the joint commitment to representing the kind as natural is partly responsible for the *appearance* that there is something much deeper to belonging to the category than meeting some agreed upon descriptive criteria. A kind is social not because the only thing that all and only members of the kind share is a social position, but rather because the kind is more explanatory than it would otherwise be, were it not for the joint commitment to representing the kind in a covertly normative way.

### *Too Cognitive Revisited*

The social positional theorist may worry that the account on offer is too cognitive. What about material conditions and practices that contribute to oppression? I do not want to say that shared intentionality provides a complete explanation of the injustices that debunkers are interested in revealing, only that it is an important part of the causal story that is obscured by standard social positional accounts. It does however seem that shared intentionality can be fruitfully understood as the vehicle of ideology and so as contributing to those very material conditions and practices that contribute to oppression.

### *Extra Ontological/Proliferation of Social Groups*

Plural subject phenomena have sometimes been critiqued as seeming extra ontological, and so one might bring into question whether it is the best sort of thing to include in a mechanistic explanation. However, to locate and describe a *mechanism* is to locate and describe entities that engage in activities that are productive of regularities.

I take the existence of plural subject phenomena to be self-evident once noticed. There is an entity that acts in accordance with certain representations. Think of a boy wearing pink to school or asking his parents for a Barbie. Even if the individuals involved do not *personally* accept the representation of boys as a natural kind, for which only certain behaviors are natural or otherwise acceptable, they are nevertheless likely to urge compliance with that representation due to the understanding that they are subject to a joint commitment and the possibility of rebuke. Insofar as the notion of plural subject phenomena enables us to explain this sort of occurrence, it appears to be a perfectly acceptable entity to appeal to in a mechanistic explanation.

It may also seem that I have replaced a proliferation of social kinds with a proliferation of social groups, since on my view there is a social group whenever there is a joint commitment to a covertly normative representation. However, again, because the parties to a joint commitment feature in mechanistic explanations, i.e., are productive of regularities among members of social kinds, I am happy to say they exist. Conversely, it has been noticed, there may be reason to be suspicious of recommending that belonging to a social kind is itself productive of regularities, i.e., is causal, as this is akin to blaming the victim.

## *Dynamic Nominalism and Radical Social Constructionism Revisited*

We avoid the critique directed at Ian Hacking, i.e., that *the point is not simply that things could have gone differently*, because the joint intentional mechanism reveals a more radical type of contingency, or at least a different type, that would be missed by focusing on mechanisms at individual or lower levels. We avoid the critique directed at the radical social constructionist, i.e., that *allowing that all kinds are dependent on us weakens the force of the social constructionist position*, because there remains a distinction between kinds that feature in regularities to a significant extent as a result of joint intentional mechanisms of a particular sort and those that do not.

## CONCLUSION

In chapter one I argued first that Haslanger draws too sharp of a distinction between natural and social kinds, and then I went on to argue that Ritchie does not make a clear enough contrast between them. My concern with Haslanger's critical realism was that it bifurcates ontology in a way that unnecessarily forces one to accept a sort of thoroughgoing realism. My concern with Ritchie's structure realization was that it seems to collapse too easily into perfectly natural structures. To avoid both outcomes simultaneously I suggested that instead of focusing on relationships of metaphysical dependence or on metaphysical explanation, we focus instead on giving a mechanistic explanation of the deeply contingent regularities of interest to debunkers, and in doing so, explaining how it is that they can appear natural, as I think this is what is crucial to capture with the notion of social construction. I have argued that Gilbertian shared intentionality is a good explanation of regularities that debunking social constructionists are interested in (deeply and in a special sense contingent) and is also not capturable in terms of individual intentions. I have argued that thus, my proposal avoids a commitment to social



essences (realist or anti-realist) while also avoiding ontological collapse. I have recommended that what makes the kinds of interest distinctively social is not that they share nothing but a social property, or that they are socially constituted (although they may well be in part), but rather that they feature in regularities due to joint commitments with particular sorts of contents. I have argued that my account therefore can capture systematicity better than the conferralist account as well as block the proliferation of social kinds. In avoiding commitment to anything like social essences I believe the account will also better capture intersectionality and passing and allow for positive identification.

## CONCLUSION

In this dissertation I examined the question of how a kind that is social can appear natural. I defended my proposal that joint intentions with certain sorts of contents are the mechanism that explain how plausibly constitutively socially constructed kinds can appear natural, and that featuring in regularities that are explained by joint intentional mechanisms is what makes a kind distinctively social in the sense of interest to the debunking social constructionist. I made a case for thinking that this proposal has some important epistemic advantages over prominent proposals for how to understand social construction claims, in particular the proposal that we understand the claim that some kind is social as the claim that the kind is constitutively socially constructed, or metaphysically dependent on social structure. These epistemic advantages include avoiding a heavyweight metaphysical view of kinds and thus accommodating a wider audience than the critical realist alone, providing a method for singling out the kinds of primary interest to the debunker, and avoiding what I have called ontological collapse, or rendering social phenomena epiphenomenal. I argued against the claim sometimes made by the debunking social constructionist that critical realism, and in particular realism about social positions, is necessary for making sense of the distinction required for social change. I made a case for thinking that critical anti-realism is a viable position and that the joint intentional mechanistic kind view, which is compatible with a sort of critical anti-realism, is better able than social positional views to capture phenomena that is important for an account of the metaphysics of social categories to capture, namely inclusion and passing, positive identification, and intersectionality.

I have argued that the thoroughgoing realism implied by social positional views and sometimes defended as necessary by the social positional theorist is not actually necessary for

successful debunking projects, and indeed is not sufficient for them. I have argued that social positional views are both over- and under- inclusive when it comes to isolating a particularly important form of social construction/type of social kind, and so leave certain epistemic benefits out of reach. As an alternative I made a case for thinking that the kinds of most interest to debunkers are those which feature in regularities in part due to the existence of a commitment to a particular ideology or an adoption of a particular conceptual schema, i.e., due to the existence of joint intentional mechanisms. What makes a kind interestingly social, a kind that debunkers ought to take particular interest in, is when our collective acceptances about what is appropriate for one kind of thing and not another plays a large role in explaining regularities among members of the kind. This is not true of all and only those kinds purported to be perceptually individuated social positions.

One important upshot of this proposal is that critical anti-realism is a viable position for the debunking social constructionist, because the crucial thing to recognize is not a special relationship of metaphysical dependence, but rather a uniquely radical type of contingency. We capture all and only the kinds that intuitively feature in the relevant type of radically contingent regularities. Debunkers can be neutral on whether there are essences. They can hold that there are natural and social essences, they can hold that all essences are conferred. What is crucial for blocking pernicious generalizations is locating and describing the mechanism that explains how kinds that are social can appear natural. Debunkers can capture the distinction between natural and social kinds without committing to critical realism, so long as they highlight the different mechanisms involved in making social kinds explanatory. Heavyweight ontological commitments like those to real social positions and everything that is implied, are not *necessary* to capture the core insight of debunking style constructionism. What is necessary is the locating

of the mechanism responsible for homeostasis in the cases of interest. Locating a mechanism does not require commitment on these heavyweight ontological issues. I suggest that instead of focusing on relationships of metaphysical dependence or on metaphysical explanation, we focus instead on giving a mechanistic explanation of the deeply contingent regularities of interest to debunkers, and in doing so, explaining how it is that they can appear natural, as I think this is what is crucial to capture with the notion of social construction.

Another upshot of this proposal is that it avoids encouraging epiphenomenalism about the phenomena of interest. Plural subjects are material objects that have intentionality. They do not straightforwardly collapse into natural structures. That is because it need not be the case that all the individuals that are jointly committed are also personally committed. I have argued that shared intentional mechanisms are a good explanation of regularities that social constructionists are interested in (deeply and in a special sense contingent) and is also not capturable in terms of individual intentions. Thus, my proposal avoids a commitment to social essences while also avoiding ontological collapse. I have recommended that what makes the kinds of interest distinctively social is not that they share nothing but a social property, or that they are socially constituted (although they may well be in part), but rather that they feature in regularities as a result of the activities of a plural subject. In avoiding social essences I believe the account will also better capture intersectionality and passing and allow for positive identification.

I put important discussions in the philosophy of social phenomena in conversation with those happening in the social construction literature. I recommended a return to understanding social construction in terms of causal explanations rather than metaphysical ones while avoiding shortcomings with existing causal accounts of social construction with my distinctive focus on the importance of joint intentional mechanisms in making sense of the significance of social

construction claims. One important final upshot of the proposal defended here is that, because jointly accepted representations explain regularities among members of social kinds, it becomes clear that a change in our collective acceptances is required for altering unjust social arrangements. Locating the mechanism that explains how regularities that appear natural, inevitable, and so do not seem to require justification allows us to not only to show the need for justification, but to reveal a method for altering unjust social arrangements. We may do so by singling out the joint intentions that perpetuate injustices for certain collections of individuals and finding ways to get uptake for the acceptance of new representations.

As Sally Haslanger (2017) has argued, we are guided in our everyday interactions by ideology. Ideology helps us to coordinate, and there is a drive for coordination. Thus, ideology seems like something we are plausibly jointly committed to as a culture. If we accept that we are jointly committed to ideology, then on Gilbert's picture, we can be rebuked for non-compliance with it. As we have seen, the parties to a joint commitment constitute a *plural subject*, and social groups are in, in a central sense of the term, plural subjects. Thus, members of a social group, if we allow that they are parties to certain joint commitments, will have standing to demand of each other behavior that is consistent with those joint commitments, however implicit the agreement that gave rise to them may be.<sup>76</sup> While not everyone may find it intuitive that such standing is legitimate, or gives rise to genuine obligation, there is little room for doubt that there is a normative expectation involved and that it shapes the behavior of group members.

---

<sup>76</sup> Gilbert writes, "plural subjects can be large, long-lasting, hierarchical, impersonal, and anonymous: the individual members need not know each other personally, or even know of each other as individuals. Certainly, they need not have sat down round a table and hammered out an agreement for all to sign" (Gilbert, 2018, p. 226).

How do we change out joint commitments? In future work I would like to explore the possibility of a phenomenon hinted at early in Gilbert's (2018) book, *Rights and Demands*, that acting as if one has standing can sometimes get one that standing. Gilbert writes,

This possibility may encourage some people to speak to others in a demanding tone when they know they lack the standing to demand the action in question. Their tendentious use of this tone may, in effect, gain them the standing they desire" (Gilbert, 2018, p. 58).

I think this idea can be fruitfully put in conversation with Rae Langton's (2017) work on the phenomenon of *accommodation* and Elizabeth Anderson's (2016) idea of a democratic form of collective moral inquiry. In her work on the phenomenon of accommodation, Langton discusses the idea that acting as if something is true can make it true, especially in conversation. This idea comes from David Lewis (1979), but Langton applies it specifically to understanding the authority of hate speech. She writes "authority can be obtained by *accommodation*, a default adjustment that occurs, without fuss, when hearers take on board what speakers presuppose" (Langton, 2017, p. 4). The common thread is that a claim made without standing (Gilbert, 2018) or a statement made without authority (Langton, 2017) can gain standing or authority with the proper uptake. Compare this common thread with Anderson's idea of a democratic form of collective moral inquiry:

"On this view, social groups learn to improve their moral norms through historical processes of contention over them... Eg. Interpersonal claims, petitions, hearings, testimonials, election campaigns, voting, bargaining, litigation, demonstrations, strikes, disobedience, rebellion...We can model the epistemic value of different modes of contention in terms of their potential for inducing error-correction, counteracting bias, clearing up confusion, taking up morally relevant information, making people receptive to admitting mistakes, drawing logical conclusions, and other epistemic improvements" (Anderson, 2016, p. 93).

Putting these ideas together it becomes clear that if the joint intentional mechanisms discussed in this dissertation can serve to encourage and sustain oppressive practices, then this seems to imply

that oppressive practices could be discouraged and prevented by these same mechanisms. As

Langton eloquently puts it:

“But hate speech can also lose strength. On this picture, the force of hate speech is partly hostage to the responses of hearers and bystanders...the unnoticed strength of hate speech...therefore goes hand in hand with an unnoticed weakness” (Langton, 2017, p. 26)

The germ of a suggestion for how we might change our joint commitments can be found in Gilbert’s gesture toward the possibility of demanding standing, in Langton’s noting of the chink in the armor of authority gained by accommodation, in Anderson’s and Haslanger’s insights about the role of resistance and disobedience in disrupting the coordination that sustains ideological oppression.

## BIBLIOGRAPHY

- Alcoff, Linda Martín (2006). *Visible Identities: Race, Gender, and the Self*, Studies in Feminist Philosophy (New York, 2006; online edn, Oxford Academic, 1 Feb. 2006)
- Andersen, Holly (2014). “A Field Guide to Mechanisms: Part I”. *Philosophy Compass* 9 (4):274-283. DOI 10.1111/phc3.12119
- Andersen, Holly (2014). “A Field Guide to Mechanisms: Part II”. *Philosophy Compass* 9 (4):284-293. DOI 10.1111/phc3.12118
- Anderson, Elizabeth (2016). “The social epistemology of morality: learning from the forgotten history of the abolition of slavery”. In Michael Brady & Miranda Fricker (eds.), *The Epistemic Life of Groups: Essays in the Epistemology of Collectives*. Oxford University Press UK.
- Andler, Matthew (2019). “Sexual Orientation, Ideology, and Philosophical Method” *Journal of Social Ontology*, vol. 5, no. 2, 2019, pp. 205-227. <https://doi.org/10.1515/jso-2019-0033>
- Andler, Matthew (Forthcoming). “Kinship and Demarcation”. Presented at Central APA 2023 in Denver
- Appiah, Anthony (1985). “The Uncompleted Argument: Du Bois and the Illusion of Race.” *Critical Inquiry*, vol. 12, no. 1, 1985, pp. 21–37.
- Appiah, Kwame Anthony (1996). *Race, Culture, Identity: Misunderstood Connections*. The Tanner Lectures on Human Values 17:51-136.
- Ásta (2008). “Essentiality conferred”. *Philosophical Studies* 140 (1):135 - 148.



- Ásta (2011). “The Metaphysics of Sex and Gender”. In Charlotte Witt (ed.), *Feminist Metaphysics*. Springer. pp. 47--65.
- Ásta (2013a). “Knowledge of essence: the conferralist story”. *Philosophical Studies* 166 (1):21-32.
- Ásta (2013b). The Social Construction of Human Kinds. *Hypatia* 28 (4):716-732.
- Ásta (2017). “Social Kinds”. *The Routledge Handbook of Collective Intentionality*, ed. by Kirk Ludwig and Marija Jankovic. Routledge.
- Ásta (2018). *Categories We Live By: The Construction of Sex, Gender, Race, and Other Social Categories*, Oxford University Press.
- Audi, Paul (2012). Grounding: Toward a Theory of the In-Virtue-Of Relation. *Journal of Philosophy* 109 (12):685-711. DOI 10.5840/jphil20121091232
- Baker, Lynne Rudder (2002). The Ontological Status of Persons. *Philosophy and Phenomenological Research* 65 (2):370-388.
- Baker, Lynne Rudder (2013). *Naturalism and the First-Person Perspective*. New York, US: Oxford University Press.
- Barnes, Elizabeth (2012). “Emergence and Fundamentality,” *Mind*, Vol. 121.
- Barnes, Elizabeth (2017). “Realism and social structure”. *Philosophical Studies* 174 (10):2417-2433.
- Barnes, Elizabeth (2018). “Symmetric Dependence”. In Ricki Leigh Bliss & Graham Priest (eds.), *Reality and Its Structure*. Oxford, UK: pp. 50-69.

- Beauvoir, Simone de (1949). *Le Deuxième Sexe*. Paris: Librairie Gallimard, 1949; *The Second Sex*. Trans. and Ed. H. M. Parshley. New York: Alfred A. Knopf, 1953.
- Bird, Alexander (2018). “The metaphysics of natural kinds”. *Synthese* 195 (4):1397-1426. DOI 10.1007/s11229-015-0833-y
- Bliss, Ricki & Trogon, Kelly (2021). “Metaphysical grounding”. *Stanford Encyclopedia of Philosophy*.
- Boyd, R. (1999a). “Homeostasis, Species, and Higher Taxa”, in R. Wilson (ed.), *Species: New Interdisciplinary Essays*, Cambridge, MA: MIT Press: 141–186.
- Boyd, R. (1999b). “Kinds, Complexity and Multiple Realization”, *Philosophical Studies*, 95: 67–98.
- Bratman, Michael (1993). “Shared Intention.” *Ethics*, vol. 104, no. 1, pp. 97–113.
- Bratman, Michael (1999). “I Intend that We J” in *Faces of Intention: Selected Essays on Intention and Agency*. Cambridge University Press. pp. 142–161.
- Burge, Tyler (1979). “Individualism and the Mental.” In *Midwest Studies in Philosophy 4: Studies in Metaphysics*, eds. P.A. French, T.E. Uehling, and H. Wettstein. Minneapolis: University of Minnesota Press, pp. 73-121
- Butler, Judith (1990). *Gender Trouble: Feminism and the Subversion of Identity*. Routledge, New York, 33.
- Butler, Judith (1993). *Bodies That Matter: On the Discursive Limits of Sex*. Routledge.
- Calhoun, Cheshire (2002). *Feminism, the Family, and the Politics of the Closet: Lesbian and Gay Displacement*, Oxford University Press.

- Callahan, Gerald N. (2009). *Between XX and XY Intersexuality and the Myth of Two Sexes*. Chicago Review Press.
- Crenshaw, Kimberle (1989). “Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics.” *University of Chicago Legal Forum* 140:139–167.
- Crenshaw, Kimberle (1991). “Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color.” *Stanford Law Review* 43, no. 6: 1241–1299.
- Dembroff, Robin (2016). “What is Sexual Orientation?” *Philosophers’ Imprint* 16(3): 1-27.
- Díaz-León, Esa (2019). “On how to achieve reference to covert social constructions”. *Studia Philosophica Estonica* 12:34-43.
- Díaz-León, Esa (2022). “Sexual Orientations: The Desire View”. In Keya Maitra & Jennifer McWeeny (eds.), *Feminist Philosophy of Mind*. New York, NY: Oxford University Press. pp. 294-310.
- Epstein, Brian (2015a). *The Ant Trap: Rebuilding the Foundations of the Social Sciences*. New York, NY: Oxford University Press.
- Epstein, Brian (2015b). “A Framework for Social Ontology”. *Philosophy of the Social Sciences* 46 (2):147-167. DOI 10.1177/0048393115613494
- Fausto-Sterling, Anne (2000a). *Sexing the Body*. New York: Basic Books, 2000.
- Fausto-Sterling, Anne (2000b). “The Five Sexes, Revisited”. *The Sciences*, July/August 2000.
- Gilbert, Margaret (1987). “Modelling collective belief”. *Synthese* 73 (1):185-204.

- Gilbert, Margaret (1989). *On Social Facts*. Princeton University Press.
- Gilbert, Margaret (1990). "Walking Together: A Paradigmatic Social Phenomenon". *Midwest Studies in Philosophy* 15 (1):1-14.
- Gilbert, Margaret (2006). *A Theory of Political Obligation: Membership, Commitment, and the Bonds of Society*. Oxford, GB: Oxford University Press.
- Gilbert, Margaret (2009). "Shared intention and personal intentions". *Philosophical Studies* 144 (1):167 - 187.
- Gilbert, Margaret (2013). *Joint Commitment: How We Make the Social World*. New York, NY: OUP USA.
- Gilbert, Margaret (2018). *Rights and Demands: A Foundational Inquiry*. Oxford, United Kingdom: Oxford University Press.
- Gilbert, Margaret (2023). *Life in Groups: How We Think, Feel, and Act Together*. Oxford, GB: Oxford University Press.
- Hacking, Ian (1999). *The Social Construction of What?* Harvard University Press.
- Haslanger, Sally (2012). *Resisting Reality: Social Construction and Social Critique*. New York, US: Oxford University Press.
- Haslanger, Sally (2016). "What is a (social) structural explanation?" *Philosophical Studies* 173 (1):113-130.
- Haslanger, Sally (2017). "Culture and Critique." *Proceedings of the Aristotelian Society* Supplementary Volume XCI (2017), pp. 149-173.

- Haslanger, Sally (2020). "Failures of Individualism: The Materiality of Social Systems." *Journal of Social Philosophy*. <https://doi.org/10.1111/josp.12373>
- Jenkins, Katharine (2016). "Amelioration and Inclusion: Gender Identity and the Concept of Woman". *Ethics* 126 (2):394-421.
- Khalidi, Muhammad Ali (2015). "Three Kinds of Social Kinds". *Philosophy and Phenomenological Research* 90 (1):96-112.
- Kim, Jaegwon (1990). "Supervenience as a philosophical concept". *Metaphilosophy* 21 (1-2):1-27.
- Kim, Jaegwon (1993). "Postscripts on supervenience". In *Supervenience and mind: Selected philosophical essays* (pp. 161–74). Cambridge: Cambridge University Press.
- Kim, Jaegwon (2006). "Emergence: Core ideas and issues". *Synthese* 151 (3):547-559.
- Knobe, Joshua & Prasada, Sandeep & Newman, George (2013). "Dual Character Concepts and the Normative Dimension of Conceptual Representation" *Cognition*. 127. 242-257.  
10.1016/j.cognition.2013.01.005.
- Kripke, Saul (1981). *Naming and Necessity*. Wiley-Blackwell.
- Langton, Rae (2017). "The Authority of Hate Speech," *Oxford Studies in Philosophy of Law*, Vol. 3.
- Leslie, Sarah-Jane (2015). "'Hillary Clinton is the Only Man in the Obama Administration': Dual Character Concepts, Generics, and Gender," *Analytic Philosophy*, 56: 111-141

- Leslie, Sarah-Jane and Adam Lerner (2022). “Generic Generalizations”, *The Stanford Encyclopedia of Philosophy* (Fall 2022 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <<https://plato.stanford.edu/archives/fall2022/entries/generics/>>.
- Lewis, David (1969). *Convention: A Philosophical Study*. Cambridge, MA, USA: Wiley-Blackwell.
- Lewis, David (1979). “Scorekeeping in a Language Game”, *Journal of Philosophical Logic* 8: 339–59.
- Machamer, Peter; Darden, Lindley & Craver, Carl F. (2000). “Thinking about mechanisms”. *Philosophy of Science* 67 (1):1-25.
- MacKinnon, Catharine (1982). “Feminism, Marxism, Method, and the State: An Agenda for Theory,” in *Feminist Theory: A Critique of Ideology*, ed. Nannerl O. Keohane, Michelle Z. Rosalvo, and Barbara C. Gelpi, University of Chicago Press, 1982: 1-30.
- MacKinnon, Catharine (1983). “Feminism, Marxism, Method, and the State: Toward Feminist Jurisprudence,” *Signs* 8 (1983): 635-58.
- MacKinnon, Catharine (1989a). *Feminism Unmodified*. Harvard University Press.
- MacKinnon, Catharine (1989b). *Towards a Feminist Theory of the State*. Harvard University Press.
- Mallon, Ron (2004). “Passing, traveling and reality: Social constructionism and the metaphysics of race”. *Noûs* 38 (4):644–673.
- Mallon, Ron (2006). “‘Race’: Normative, Not Metaphysical or Semantic”. *Ethics* 116 (3):525-551.

- Mallon, Ron (2007a). “A field guide to social construction”. *Philosophy Compass* 2 (1):93–108.
- Mallon, Ron (2007b). “Human categories beyond non-essentialism”. *Journal of Political Philosophy* 15 (2):146–168.
- Mallon, Ron (2008). “Naturalistic Approaches to Social Construction”, *The Stanford Encyclopedia of Philosophy* (Spring 2019 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2019/entries/social-construction-naturalistic/>.
- Mallon, Ron (2013). “Naturalistic approaches to social construction”. *Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/entries/social-construction-naturalistic/>.
- Mallon, Ron (2016). *The Construction of Human Kinds*. Oxford University Press.
- Manley, David (2009). “Introduction” in Wasserman, Ryan; Manley, David & Chalmers, David (eds.) (2009). *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford, England: Oxford University Press.
- Mason, Rebecca (2016). “The metaphysics of social kinds”. *Philosophy Compass* 11 (12):841-850.
- Mason, Rebecca (2021). “Social kinds are essentially mind-dependent”. *Philosophical Studies* 178 (12):3975-3994.
- Mill, John Stuart (1843). *A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence, and the Methods of Scientific Investigation*. London, England: Longmans, Green, Reader, and Dyer.
- Mills, Charles W. (1998). *Blackness Visible: Essays on Philosophy and Race*. Cornell University Press.

O'Connor, Cailin (2019). *The Origins of Unfairness: Social Categories and Cultural Evolution*.  
Oxford University Press.

Pettigrove, Glen. (2016). "Changing Our Mind," *The Epistemic Life of Groups: Essays in the Epistemology of Collectives*, Ed. by Michael S. Brady and Miranda Fricker. Oxford University Press, New York.

Putnam, Hilary (1973). "Meaning and Reference." *The Journal of Philosophy* 70: 699-711

Putnam, Hilary (1975). "The Meaning of 'Meaning'." In *Mind, Language and Reality*.  
Cambridge: Cambridge University Press, pp. 215-71.

Quine, Willard V. (1969). "Natural kinds". In Jaegwon Kim & Ernest Sosa (eds.), *Ontological Relativity and Other Essays*. Columbia University Press. pp. 114-38.

Richardson, Kevin (Forthcoming). "The Scalar Theory of Social Categories".

Ritchie, Katherine (2018). "Social Creationism and Social Groups". *Collectivity: Ontology, Ethics, and Social Justice*. (K. Hess, T. Isaacs, and V. Ighneski (Eds.), Rowman & Littlefield)

Ritchie, K. (2019), "Should We Use Racial and Gender Generics?", *Thought: A Journal of Philosophy*, 8(1): 33-41

Ritchie, Katherine (2020) "Social Structures and the Ontology of Social Groups". *Philosophy and Phenomenological Research* 100 (2):402-424.

Ritchie, Katherine (2021). "Does Identity Politics Reinforce Oppression?" *Philosophers' Imprint* 21 (4):1-15.



- Rodriguez-Pereyra, Gonzalo (2008). "Nominalism in Metaphysics." *The Stanford Encyclopedia of Philosophy*. Ed. E. N. Zalta.  
<http://plato.stanford.edu/archives/fall2008/entries/nominalism-metaphysics>.
- Ross, L. N. (2023). What is social structural explanation? A causal account. *Noûs*, 00, 1– 17.
- Roughgarden, J. (2004) *Evolution's Rainbow: Diversity, Gender and Sexuality in Nature and People*. University of California Press, Berkeley and Los Angeles.
- Salmon, Wesley (1984). *Scientific Explanation and the Causal Structure of the World*, Princeton University Press.
- Salmon, Wesley (1990). *Four Decades of Scientific Explanation*. University of Pittsburgh Press.
- Schaffer, Jonathan (2009). "On what grounds what". In D. Chalmers, D. Manley, & R. Wasserman (Eds.), *Metametaphysics: New essays on the foundations of ontology* (pp. 347–383). Oxford: Oxford University Press.
- Schaffer, Jonathan (2012). "Grounding, transitivity, and contrastivity". In Fabrice Correia & Benjamin Schnieder (eds.), *Metaphysical Grounding: Understanding the Structure of Reality*. Cambridge University Press. pp. 122-138.
- Schaffer, Jonathan (2017). "Social construction as grounding; or: fundamentality for feminists, a reply to Barnes and Mikkola". *Philosophical Studies* 174 (10):2449-2465.
- Schweikard, David P. and Hans Bernhard Schmid (2021). "Collective Intentionality", *The Stanford Encyclopedia of Philosophy* (Fall 2021 Edition), Edward N. Zalta (ed.), URL = [<https://plato.stanford.edu/archives/fall2021/entries/collective-intentionality/>](https://plato.stanford.edu/archives/fall2021/entries/collective-intentionality/).

- Searle, John (1990). "Collective Intentions and Actions". In Philip R. Cohen Jerry Morgan & Martha Pollack (eds.), *Intentions in Communication*. MIT Press. pp. 401-415.
- Searle, John R (1995). *The Construction of Social Reality*. Free Press.
- Stein, Edward (1999). *The Mismeasure of Desire: The Science, Theory and Ethics of Sexual Orientations*, Oxford University Press.
- Stoljar, Natalie (2011). "Different Women. Gender and the Realism-Nominalism Debate" in *Feminist Metaphysics*. Ed. Charlotte Witt.
- Tuomela, Raimo and Kaarlo Miller (1988). "We-Intentions," *Philosophical Studies*, 53: 367–89.
- Tuomela, Raimo (2007). *The Philosophy of Sociality: The Shared Point of View*. New York, US: Oup Usa.
- Van Fraassen Bas, C. (1980). *The Scientific Image*. Oxford, England: Oxford University Press.
- Wilkerson, William (2013). "What is 'Sexual Orientation'?" in Powell, Halwani & Soble (eds.) *Philosophy of Sex*, 6<sup>th</sup> edition, Rowman & Littlefield.
- Woodward, James (2003). *Making Things Happen: A Theory of Causal Explanation*. New York, US: Oxford University Press.
- Woodward, James and Ross, L. N. (2021). "Scientific Explanation", *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition), Edward N. Zalta (ed.), URL = [<https://plato.stanford.edu/archives/sum2021/entries/scientific-explanation/>](https://plato.stanford.edu/archives/sum2021/entries/scientific-explanation/).
- Zack, Naomi (1993). *Race and Mixed Race*. Temple University Press.
- Zack, Naomi (2002). *Philosophy of Science and Race*. New York: Routledge.