# Barriers to Change, Possibilities for Resistance

Concepts within Structures of Oppression, Obstacles to Innovation, and the Implementation Challenge of Conceptual Engineering

**Paul-Mikhail Catapang Podosky**
ORCID: 0000-0001-5636416X
Doctor of Philosophy
12th February, 2021

School of Historical and Philosophical Studies, University of Melbourne
A thesis submitted in total fulfilment for the degree of Doctor of Philosophy

# Abstract

Conceptual engineering, when it comes to social kind concepts, has strong political roots within the academy and activist circles alike. But if conceptual engineering, understood as the development of non-dominant conceptual practices, is to be a useful tool for the purposes of achieving or contributing to social justice, there must be a means by which the concepts we design in theory, or within small communities of practice, can take root and propagate in dominant contexts. This, broadly speaking, is known as the implementation challenge of conceptual engineering.

This thesis has the general task: to reorient how we approach the implementation challenge. It makes explicit and criticises existing accounts of conceptual implementation, namely those that focus on the role of individualistic implementation strategies in bringing about conceptual change within a community. This emphasis on individual conceptual advocacy warps our perception of the shape, size, and nature of the problem. It fails to recognise our situatedness in social structures that work to maintain and entrench the conceptual status quo, and which stifle conceptual innovation. In this thesis, I identify two mechanisms of conceptual maintenance: *psychological convergence mechanisms* and *mechanisms of conceptual reproduction*. The former refers to the social processes by which thinkers and speakers gain similar knowledge structures, in particular overlapping characterizations (i.e. stereotypes), within a community; and the latter, which will be the core focus of this thesis, refers to the mechanisms that promote the copying of prior stable patterns of classification with a term. Both mechanisms are ubiquitous within our social and representational milieu, operating within formal institutions to everyday conversation and engagement with social reality, and working to preserve dominant terms of conceptual engagement. Importantly, both reduce the likelihood of an individual successfully motivating others, within dominant contexts, to adopt an engineered conceptual practice.

After general discussion of mechanisms of conceptual maintenance, I spend time explicating certain problems for a particular individualistic implementation strategy that focuses on the reproduction of alternative conceptual practices within in interpersonal speech situations. In particular, the strategy involves engaging with dominant speakers in conversation as to which concept should be expressed by a shared word in a context (i.e. metalinguistic disagreement). The hope is that the dominant speaker will recognise that an extant conceptual practice is deficient, or stands in need of improvement. I argue that such forms of disagreement are often infected with unjust power relations that tend to advantage dominant speakers and existing patterns of classification. This contributes to preserving the conceptual status quo, and subsequently suppresses conceptual change. Moreover, when metalinguistic disagreement favours dominant speakers, I argue that this often constitutes hitherto undiscussed forms of epistemic and linguistic injustice. Overall, my aim is to show that when we develop implementation strategies, we should be careful to take into account the social infrastructure and forces that work to keep things conceptually as they are.

# Declaration

With the exception of the sections listed in the preface, the work in this thesis is original and completely my own. Due acknowledgement has been made in this thesis to all other material that has been used. This thesis is fewer than 100,000 words. No part of this thesis has been submitted for any degree or is currently being submitted for any other degree.

Paul-Mikhail Catapang Podosky
12th of February 2021

# Preface

Chapter one, section 1.2 of this thesis draws on parts of a co-authored paper with Kai Tanter. Our contributions were equal.

Further, parts of this thesis have been published during the process of completion. The publication details for each article are as follows:

1. Podosky, Paul-Mikhail Catapang. 2021. Agency, Power, and Injustice in Metalinguistic Disagreement. *The Philosophical Quarterly*. pp. 1 = 30. DOI: 10.1093/pq/pqab023
   **Accepted:**

2. Podosky, Paul-Mikhail Catapang. 2021. Rethinking Epistemic Appropriation. *Episteme: A Journal of Individual and Social Epistemology*. pp. 1 – 21. DOI: 10.1017/epi/2021.8
   **Accepted:** 29/01/2021

3. Podosky, Paul-Mikhail Catapang. 2020. Gaslighting, First and Second Order. *Hypatia: A Journal of Feminist Philosophy*. pp. 1 – 21. DOI: 10.1017/hyp2020.54.
   **Accepted:** 19/04/2020

4. Podosky, Paul-Mikhail Catapang and Kai Tanter. 2020. Revision, Endorsement, and the Analysis of Meaning. *Analysis* 80 (4): pp. 1 – 15. DOI: 10.1093/analys/anaa010.
   **Accepted:** 18/02/2020.

5. Podosky, Paul-Mikhail Catapang. 2019. Privileged Groups and Obligation: Engineering Bad Concepts. *Journal of Applied Philosophy*. pp. 1 – 16. DOI: 10.1111/japp.12398.
   **Accepted:** 18/09/2019.

6. Podosky, Paul-Mikhail Catapang. 2018. Ideology and Normativity: Constraints on Conceptual Engineering. *Inquiry*: *An Interdisciplinary Journal of Philosophy*. pp. 1 – 15. DOI: 10.1080/0020174X.2018.1562374.
   **Accepted:** 14/12/2018

7. Podosky, Paul-Mikhail Catapang. 2018. 'What Defines a Conceptual Resource?' *Ergo: An Open Access Journal of Philosophy* 6 (6): 143 – 167. DOI: 10.3998/ergo.12405314.0006.006.
   **Accepted:** 13/08/2018

# Funding

I would like to acknowledge that this thesis was completed with the help of the Australian Government Research Training Program Scholarship.

# Acknowledgements

My father had serious heart complications this year. And for a moment I thought that I had lost him. But through his own unrelenting effort, he was able to overcome this life hurdle. I am grateful for his strength. For my father to see the completion of this thesis truly means the world to me. He is the reason that I do philosophy, and the reason why have been able to excel. Thank you, Michael Podosky, PhD.

To Madison Griffiths, who has been with me throughout the development of this thesis: I am lucky to have you in my life. I know I was hard to be around at times. Thank you for the insightful conversations over the past four years. Each contributed to ensuring that the ideas and arguments in this thesis stayed true to real world experiences.

I could not have asked for better supervisors than Associate Professor Karen Jones and Dr Laura Schroeter. The guidance of both made for the perfect intersection of ideas. This thesis would not have been possible without their wisdom, care, and endless support. I am filled with gratitude.

There are many friends that I have exchanged thoughts with throughout my candidature. Thank you for being there for me. I am especially grateful to the philosophy postgraduate cohort and the Minorities and Philosophy (MAP) group at the University of Melbourne. I will sorely miss my role as the Colloquium Officer.

Finally, I would like to acknowledge that this thesis was primarily written on the lands of the Wurundjeri People of the Kulin Nation. Sovereignty has never been ceded. It always was and always will be Aboriginal Land.

# Contents

# Introduction

## Why Concepts Matter

*Arguably, our conceptual repertoire determines not only what beliefs we can have but also what hypotheses we can entertain, what desires we can form, what plans we can make on the basis of such mental states, and accordingly constrain what we can hope to accomplish in the world.*

Alexis Burgess and David Plunkett 2013, p. 1097

### I. Milk and Plants

In September of 2019, The Nationals, a conservative political party in Australia, set to lobby the Federal Government to prohibit the use of the term 'milk' from being used for plant-based products such as oat, soy, and almond milk.[1] This push for reform followed a global trend. For example, Canada has seen the introduction of laws to ban the use of dairy alternatives being labelled as 'milk.' In the U.S., there is bipartisan support for the DAIRY PRIDE Act, which would put pressure on the Food and Drug Administration to offer a precise definition of 'milk.' And in Europe, the Court of Justice has ruled that purely plant-based products cannot, in principle, be marketed with designations such as 'milk', 'cream', 'butter', 'cheese' or 'yoghurt.'

One might be baffled by this effort: What's in a word, after all? The answer is not so simple. It's not *just* a word that is the problem.[2] Instead, it is what a word *expresses.* And what a word expresses is a *concept.* The disagreement, then, is over which concept a word *should* express: which concept should be expressed by 'milk'? Before I explain the nature of such normative conceptual questions in more detail, let's explore some of the reasons why there has been a push for legal reform to prohibit the use of dairy terms for plant-based items.

Preceding The National Party's lobbying of the Federal Government, dairy farmers in Australia faced one of the toughest years on record. Owing to bad weather, lack of government support, and consumers choosing to buy alternatives to dairy, dairy-based milk production plummeted to a 21 year low. This translated into a significant financial loss for dairy farmers. Unfortunately for farmers and lobbyists, the weather is not something that can be changed – at least not easily. However, there is a certain degree of control to be had over consumers. And this can be achieved through government intervention. Herein lies the reason for placing a legal ban on the use of words: if dairy milk sales are affected by consumers choosing to buy other types of milk, then there is at least *prima facie* reason to think that consumers will not choose to buy other types of milk *if no such types exist.* This does not mean banning the sale of plant-based alternatives. Instead, it involves a normative conceptual move. It is about restricting the concept that is expressed by 'milk' to exclude anything other than the milk that comes from cows (or mammals). Thus, if the word 'milk' is to be used *only* for dairy milk, this means

[1] As well as lobbying against the use of other terms such as 'meat' and 'seafood' to apply to things such as tofu sausages and seitan prawns.

[2] Words are *part* of the problem. Herman Cappelen (2018) talks about the 'lexical effects' of words, which might give us reasons to keep the same word to express a different concept.

that other options won't factor into consumers' decisions as to which *milk* to buy. After all, no other options will exist. In other words, it is *prudential* and *political* reasons that have motivated lobbyists to push for legal reform.

On the other side of this debate, animal welfare advocates have endorsed this natural movement in our language. They suggest that the more expansive concept expressed by 'milk' is a good thing. However, the conceptual shift to include plant-based alternatives is not the end goal. Animal advocacy groups such as *PETA* have encouraged the use of terms such as 'milk' to apply to plant-based alternatives and *only* plant-based alternatives.[3] This isn't to say that organizations such as *PETA* believe that the milk that comes from cows isn't milk. It is only to say that they believe that dairy milk is not milk for human consumption. This is part of a broader aim to shift the public's understanding of the kinds of things that count as 'food.' The idea is that once we see that ideology has masked the moral significance of non-human animals, we will come to realize that the reasons for why we do not count *humans* as food are precisely the same reasons for why we should not count *other* animals as food. In other words, it is *moral, social,* and *political* reasons that have guided animal advocates to change our understanding of what should be expressed by food terms.

There are two things worth highlighting in this example. The first is this: concepts are interrelated. The use of words, that express concepts, bring into play other concepts through inferential connections. Both sides of the disagreement recognise, to some extent, that when the concept MILK is activated, other concepts are primed for activation, such as COFFEE. Put crudely: on one side of the debate, the goal is to ensure, for prudential and political reasons, that only dairy milk has a secure inferential connection to coffee; on the other side of the debate, the goal is to ensure, for moral, social, and political reasons, that coffee is only connected to plant-based items.

In an article for *The Guardian*, Lynne Tirrell makes explicit the relationship between words and concepts in her discussion of Donald Trump's use of 'infestation' for people coming to the U.S. at the southern boarder:

> [O]ur words don't actually work just one-by-one. Each word is part of a network of related terms that give it conceptual force. If someone calls a group "an infestation", hearers can immediately infer that the group being discussed is unwanted, dangerous or harmful. We talk about *infestation* of insects or rodents, so the very use of this term bring these dehumanizing concepts into play (Tirrell 2019).

Tirrell warns that we should take lesson from history. The use of 'invaders' by Rwandan political leader Leon Mugesera led to mass genocidal mobilization against the Tutsis. This takes us to the second important fact: our means of representing and classifying the world, through interconnected words and concepts, shape social reality. Representation and classification are inextricable parts of our social practices. And the term 'milk' is a clear example of this. The concept expressed by this word will influence our decisions as to what we will consume, and therefore reinforce the structures that make such consumption possible.

The disagreement over which concept should be expressed by 'milk' is not unique. We can see the structure of this back-and-forth in disagreements over other socially and politically contested terms. Think of highly institutionalized terms such as 'marriage' that required legal reform in order to express a different or revised concept, and less formal terms that have been conceptually contested such as 'gaslighting,' 'sexual harassment,' 'abuse,' and the like. As was said above, there is an intimate

[3] And, of course, human-sourced milk *for babies.*

connection between concepts and the social world. The concepts expressed by shared terms shape the direction of our engagement with our surroundings, and with those we share our surroundings with. The formation of the social world owes much to the way that we conceive of the things around us. Without concepts, social activity would be severely limited. One might now start to see why the effort to take control of the concepts expressed by words is not so baffling, after all.

## II. Concepts, in a Little More Detail

To see a little more clearly how concepts are related to the social world, first consider the role of concepts in our *mental* lives. Concepts fix the thoughts that we can have (Burgess and Plunkett 2013). What is available for me to believe depends, in part, on the world presenting itself through a certain concept. I believe that there are dogs, in part, because I possess the concept DOG that enables me to pick them out. And the role of concepts in our mental lives is more than doxastic. Concepts play a sophisticated cognitive role in the formation of desire and prediction, judgment and planning. Quite apart from this cognitive role, concepts also serve to constrain affective understanding – and even shape affective states themselves. Whether one is in love, arguably, depends on taking oneself to fall under that description; love depends on a concept of LOVE (Jones 2008; Kirsch 2019).[4]

Given the role that concepts play in our cognitive and affective understanding, it's no great leap to see how concepts serve to structure our social interactions. An oft-cited passage from Burgess and Plunkett makes this connection explicit:

> Arguably, our conceptual repertoire determines not only what beliefs we can have but also what hypotheses we can entertain, what desires we can form, what plans we can make on the basis of such mental states, and accordingly constrain what we can hope to accomplish in the world. *Representation enables action*, from the most sophisticated scientific research, to the most mundane household task. It influences our options within social/political institutions, and even helps determine which institutions are so much as thinkable. *Our social roles, in turn, help determine what kinds of people we can be, what sorts of lives we can lead* (2013, p. 1097 my emphasis).

Here, the claim is this: concepts shape our mental lives, and this has consequences for how we understand ourselves and others. Concepts frame expectation, shape institutions, and constrain (our understanding of) moral, social, and political possibility. In other words, concepts render intelligible information taken from our surroundings, serving as the basis of practical activity. We interpret the world on basis of the concepts available to us, and we act in response to the world in accordance with this interpretation.

Given that concepts play such a significant role in the formation of the social world, they are also strongly implicated in the creation of *oppressive* social relations. Oppression, as I will understand it, is a group-level phenomenon (Frye 1983). It is a function of a system of interdependent behaviour and social institutions that unjustly limits life opportunities for particular classes of individuals. Concepts are *ideological* when they motivate and undergird oppressive practices. Hence, to combat oppression, we must 'break the spell' of ideology, as it were. Part of what is required to achieve this is revising the concepts that are responsible for oppressive practices; either in the form of amending or removing defective incumbent concepts, or else introducing new concepts that undergird behaviour

[4] Perhaps, more specifically, *romantic* love.

that corrects unjust relations and ends unjust resource distributions. In other words, we need to *identify* and (*re-*)*engineer* bad concepts. This is an on-going project in the philosophical literature, and for activists who are determined to bring about positive social change.

## III. Aims and Goals

Identifying and (re)engineering ideological concepts is no easy task. In the first place, we are often unaware of the concepts that are responsible for shaping oppressive social practices. But perhaps more pressingly, our plans to *implement* engineered concepts must be able to compete with the *conservative social mechanisms that maintain (ideological) conceptual practices within a thinking and speaking community*. For if we are unable to develop strategies that can work with or against the social infrastructure and forces that preserve the conceptual status quo, then the concepts that we construct in theory, and those we develop within communities, for the purposes of achieving social justice, will be of little use to us. With this in mind, this thesis has the general aim: *to shift how we approach the implementation challenge of conceptual engineering.* Before I say more about this, I will briefly explore the nature of the implementation challenge.

Roughly, the implementation challenge of conceptual engineering concerns how it is that the ameliorative strategies that we develop in theory, designed concepts to meet particular aims and goals, can be implemented out in the wild. For example, if we develop a concept of woman that is useful for analysing oppression (e.g., Haslanger 2000; Jenkins 2016), the implementation challenge asks how it is we can get people to adopt this concept and use it in their day-to-day lives, or for theoretical feminist enquiry (as opposed to the ordinary concept of woman that presumes biological essentialism).

How to understand what it means to 'implement' a concept is varied. One might aim to change the meanings (intensions and extensions) of public language expressions (e.g. Cappelen 2018), or aim to change what is expressed by shared words in a context (e.g. Sterken 2020), or aim to introduce new speaker meanings via stipulation (e.g. Pinder forthcoming; cf. Deutsch 2020), etc.[5] Given that my interest is in advancements toward the correction of unjust social structures, via the amelioration of our conceptual repertoire, I will understand the scope of the implementation challenge to range over the concepts that are shared within a community of thinkers and speakers, consisting of many social sub-groups. However, concepts, under my way of thinking, are not the meanings of terms. Rather, they are complex cognitive structures, consisting of bodies of information and dispositions (e.g., Machery 2009). In particular, such structures include socially shared characterizations (i.e. stereotypes) and community-wide patterns of classification (see **chapter one**).

To date, there has not been a comprehensive account of what constitutes the challenge of implementing an ameliorative strategy. And a failure to properly understand the implementation challenge is a setback in social justice theorizing. If overcoming oppressive social practices requires, at least in part, that we target and change ideological concepts, then we must be sensitive to the means by which ameliorative strategies can be feasibly implemented. Given this, in this thesis I will provide an alternative take on the challenge of implementing ameliorative strategies; one that deviates from prominent proposals in the current literature. I will not offer a 'solution' to the challenge. This goes beyond the scope of the thesis. Rather, I will explore a means of reorienting our perception of the problem, focusing primarily on the social obstacles that stabilize dominant terms of conceptual engagement, and which stifle conceptual innovation.

[5] See Jorem (forthcoming) for an extended discussion of these options

In recent time, the implementation challenge is presented in a way that warps our impression of the shape, size, and nature of the problem. For example, Herman Cappelen and David Plunkett claim the following:

> Once you have settled on an ameliorative strategy, you might want to do some work to implement it, that is, you might want to engage in a bit of activism on behalf of your ameliorative strategy. If that's something you want to do, it raises an 'implementation challenge': how are ameliorative strategies best implemented? (2020, p. 3). [6]

And, Alexis Burgess and David Plunkett offer something similar:

> There is what we can call attempts at conceptual implementation. Roughly, this involves a kind of advocacy, in which one tries to get some people (ranging from a long individual to a large population), to actually take up and use the concepts one is in favor of (2020, p. 5).

While we cannot infer a fully-fledged theory of implementation from these passages, what we can read from them is that the challenge of implementing an ameliorative proposal requires that one undertake some kind of activism or advocacy on behalf of their preferred concept. That is, the implementation challenge amounts to what it is one can do to persuade others in their community to adopt an alternative conceptual practice.[7]

The issue, however, is that understanding strategies for implementation in this way is far too *individualistic*. It makes it seem as if implementing an ameliorative strategy is a 'lone-wolf' endeavour. Of course, this might be perfectly fine if one understands the notion of implementation to be a matter of stipulating new meanings in a context, or modulating existing meanings to be expressed by shared words in conversation (e.g. Pinder forthcoming). However, for those interested in changing the concepts of an entire community of thinkers and speakers, focusing on what individuals can do to bring about conceptual change should be cause for concern. After all, if all that can be done to bring about conceptual change is for one to persuade and change the minds of others, then it should come as no surprise that implementing an ameliorative strategy will be a daunting and seemingly insurmountable task.

In this thesis, I argue that individualistic implementation strategies, for social kinds concepts, will often fail to compete with the social forces that tend to favour dominant terms of conceptual engagement, and which suppress conceptual change. Individuals will constantly bump up against power-infected structural obstacles that severely reduce the likelihood of convincing people in dominant contexts to adopt an ameliorated conceptual practice. Importantly, this includes

---

[6] Mark Richard calls this an *A-project*: 'a project of offering and trying to get others to accept a revisionary ameliorative analysis of a concept…' (2020, p. 360).

[7] The implementation challenge here is formulated in a way that is intended to be neutral with respect to different ways of thinking about what concepts are, and what the relevant change must be. Max Deustch (2020), however, gives a more in depth analysis of the implementation challenge by framing the problem specifically as one about stipulative semantic meanings or revisions. As said, I am not so interested in concepts understood as meanings (intensions and extensions). Subsequently, I am not so interested in meaning change. Thus, I will not discuss Deustch's position in much detail throughout the thesis. What is interesting to note about Deutsch's view is that his worry centres on our not-knowing of what it would take to implement semantic change beyond speaker intention. One way of understanding the goal of this thesis is to bring to light facts about our social structure that prevents conceptual innovation. Though, and to reiterate, I am not interested in what is stifles meaning change.

mechanisms that regulate similarity of associative thought (i.e., stereotypes) amongst group members, or what I call *psychological convergence mechanism;* and, additionally, *mechanisms of conceptual reproduction*. The latter, mechanisms of conceptual reproduction, will be the core focus of this thesis. Briefly, conceptual reproduction refers to a social process by which certain classificatory practices associated with a term are sustained through the non-accidental copying of earlier instances of that practice within a community. And *mechanisms* of conceptual reproduction are the social infrastructure that stabilize patterns of copying, or which push a classificatory practice forward. A clear example of a strong mechanism or incentive for copying prior stable patterns of classification is legal punishment and benefits which require paying attention, or accepting, certain legal categories (e.g., *marriage*, *undocumented migrant*). Importantly, I will argue that the presence of these aspects of our social structure, the role of which is to regulate our thinking and speaking practices, makes individuals, on their own, mostly powerless against a conservative machine designed to keep things as they conceptually are.

To be more specific, I will mount an argument against a particular individualistic implementation strategy. Namely, a strategy that focuses on the role of individual conceptual advocacy in speech situations where there is disagreement over which concept should be used in a context, or what the standards for the use of a term ought to be. This is often referred to as *metalinguistic disagreement.* I will argue that there are conceptual reproductive mechanisms within conversations of this kind that tend to favour dominant patterns of classification. This casts a large shadow of doubt over the plausibility of metalinguistic discourse as a location for conceptual innovation and disruption.

Additionally, I will show that when attempting to overcome these mechanisms of conceptual maintenance in interpersonal settings, speakers can be subject to a variety of distinctive forms of epistemic and linguistic injustice; specifically when someone from a marginalized group attempts to convince a dominant hearer to adopt a conceptual practice that has been developed within non-dominant situatedness. In such cases, power unfairly affects the distribution of control in the joint activity of pairing concepts with words in a way that tends to advantage dominant terms of conceptual engagement; and this can result in gaslighting, silencing, and the undermining of communicative agency. This, I will argue, gives us further reason to doubt the plausibility of individualistic strategies of conceptual implementation, such as individual conceptual advocacy in metalinguistic disagreement; especially if one is a member of a marginalized group.

I do not intend to argue that we should completely disregard the role of individualistic implementation strategies in bringing about conceptual change. There is a time and place for individuals to plant the seeds of a new conceptual practice in dominant contexts in the hope that it will take root and propagate far and wide; and this will include activities such as engaging in difficult conversation (e.g. calling out the use of trans* exclusive language). Rather, my aim is to bring to light just how hard the implementation challenge is, at least for social kind concepts, by paying attention to the social forces that maintain the conceptual status quo, and which hold back conceptual progress. I believe that we should accept that there are many different ways of responding to the implementation challenge. However, we should make sure to design strategies that will be able to compete with prevailing social structures.

## IV. Overview

In **chapter one**, I discuss conceptual engineering. Specifically, I draw a distinction between what conceptual engineering is about and what it supposed to achieve; and I spend time answering the first

half of this distinction. I set myself apart from existing accounts of conceptual engineering that focus on concepts understood as the meanings (intension and extensions) of public language expressions, and instead treat concepts as systematically linked networks of individual knowledge structures associated with, and which underlie, particular classificatory dispositions. I also discuss one important social mechanism that regulates or stabilizes similarity of knowledge structure, and thus our classificatory dispositions, across individuals with the same community: *psychological convergence mechanisms.*

In **chapter two**, I discuss the second half of the foregoing distinction. I explore why conceptual engineering is taken to be an important form of philosophical theorising. For my purposes, I argue that classificatory dispositions play a substantive role in the creation and maintenance of social reality. And given this, revising our classificatory dispositions can help promote or advance social justice. Put differently, conceptual engineering, under my way of thinking, is supposed to contribute to achieving a world in which there are more just social relations.

In **chapter three**, I discuss the implementation challenge of conceptual engineering and conceptual reproduction. I argue that implementation challenge is often posed too *individualistically*. It is framed in a way that focuses too much on the role of *individuals* in persuading or convincing others in their community to take up an alternative conceptual practice. Such strategies ignore or fail to appropriately take into account the social infrastructure that is (partly) responsible for maintaining the conceptual status quo, or which prevent individuals in the grips of an ideology from responding positively to ameliorative proposals that would promote justice. Put differently, I make explicit the entrenched structural systems that stifle conceptual change; the social forces that keep things conceptually as they are. Specifically, I discuss a particular social process that regulates community-wide classificatory dispositions: *conceptual reproduction.* Roughly, this refers to the social processes by which individuals non-accidentally copy prior stable patterns of classification within a social and representational milieu. I show that there are different social mechanisms and incentives, from formal institutions to everyday conversation, that play a substantive role in pushing a conceptual practice forward through reproduction.

In **chapter four**, I provide a general overview of a distinctive mechanism of conceptual reproduction that occurs within metalinguistic discourse called *conceptual accommodation.* This occurs when a conceptual attitude of a speaker becomes a part of the shared information that governs the permissibility of conversational moves; specifically, moves that express a particular concept in a context. I offer a quasi-formal framework that introduces two forms of conceptual accommodation, *Establishment* and *Negotiation*. The introduction of this taxonomy sets up discussion in the following chapters (five through nine) where I mount an argument against the plausibility of a particular individualistic implementation strategy that focuses on making interventions within metalinguistic discourse.

In **chapter five,** I discuss the phenomenon on *non-ideal metalinguistic disagreement.* Such disagreements occur when there is an inequality in the joint activity of pairing concepts with words. And in such cases, conceptual accommodation occurs as a result of an unequal distribution of power in the joint activity of pairing concepts with words that tends to favour or advantage dominant patterns of classification. Because such metalinguistic disagreements tend to favour or advantage dominant patterns of classification, making successful interventions within metalinguistic disagreement, as someone who endorses justice-promoting concepts, is unlikely. More often than not will one fail to get another to adopt an ameliorative proposal, or downgrade their confidence in an existing word-concept pair. The social forces operative with metalinguistic discourse that work to

maintain the conceptual status quo casts a large shadow of doubt over the friendliness of metalinguistic discourse as site at which a non-dominant conceptual practice can take root in dominant contexts.

In **chapter six**, I follow on from the previous chapter. I explore how non-ideal metalinguistic disagreements not only tend to work in favour of dominant patterns of classification, but also occur under conditions of injustice. That is, I will introduce and examine a distinctive wrong that occurs in cases of metalinguistic disagreement in which there is an *unfair* or *unjust* distribution of metalinguistic control owing to the operation of identity prejudice in a context. Such prejudices have the effect of delimiting the possible moves one can make in the joint activity of pairing concepts with words. I call this *metalinguistic injustice.* In particular, I argue that metalinguistic injustice can take the form of an epistemic injustice, a linguistic injustice, or both.

In **chapter seven**, I examine how metalinguistic injustice helps us to locate a distinctive form of *gaslighting*. Gaslighting is one means by which extant ideological concepts are reproduced in metalinguistic discourse. I draw a distinction between *first order* and *second order* gaslighting, where my primary interest is in the latter. Second order gaslighting occurs when there is disagreement over which concept should be used in a context, and where the use of words by a speaker is apt to cause a hearer to doubt her interpretive abilities *in virtue of* doubting the accuracy of her concept. I argue that this constitutes three distinctive forms of epistemic injustice: metalinguistic injustice, conceptual obscuration, and perspectival subversion.

In **chapter eight**, the final chapter, I continue discussion on the distinctive forms of injustice that occur within metalinguistic discourse by showing that conceptual accommodation, both *Establishment* and *Negotiation,* can lead to the creation of an *unjust conversational relationship*. Unlike the previous chapters, discussion of this form of injustice shows that conceptual accommodation has pernicious effects for conversation *beyond* metalinguistic discourse. Specifically, an unjust conversational relationship is instituted in a context when a speaker is unfairly assigned a subordinate discourse role, and is thus governed by unjust discourse rules, in conversation owing to the operation of identity prejudice in the context.

## V. Apologies

I have tried as much as possible in this thesis to view social injustice through an intersectional lens. Despite this, there will be parts that fail to be wholly intersectional, focusing specifically on one axis of oppression that perhaps does not capture the nature of certain forms of injustice in their entirety. Nevertheless, I hope that what I have to say still gets to the heart of many social justice issues, and offers new ways of thinking about disruption and liberation via the amelioration of social concepts.

# Chapter One

## Conceptual Engineering: What is it About?

*The word 'concept' is used in various ways; its sense is sometimes psychological, sometimes logical, and sometimes perhaps a confused mixture of both.*

Gottlob Frege 1892, p. 42

### 1.1. Introduction

When thinking about the nature of conceptual engineering, we should be careful to distinguish between two things: what conceptual engineering is *about* and what conceptual engineering is supposed to *achieve*. It is easy to confuse these two ways of thinking. After all, one might reason: shouldn't conceptual engineering be about what it is supposed to achieve? Not quite. In a rather obvious sense, conceptual engineering is about the engineering of concepts – this seems plain and simple.[8] But what conceptual engineering is supposed to achieve, at least typically, is not just conceptual change *simpliciter*. Rather, the aim is for conceptual change to achieve a specific *goal* (Burgess and Plunkett 2013; Koch forthcoming). And the goal in question varies according to the interests of theorists. Some might be motivated to change concepts with theoretical ends in mind; some might be more concerned with enacting real world consequences for social facts.

In this chapter, I will discuss the first half of the foregoing distinction, that is, I will explore different answers to the question of what conceptual engineering is about. Importantly, I will draw a general distinction between *the semantic approach* and *the psychological approach* – of course, granting that meaning and mental/linguistic dispositions can come apart. The former, broadly speaking, understands concepts as mental representations that have semantic properties, such as public meanings or semantic values. The latter, in contrast, is about psychological information structures and linguistic behaviour or competencies such as dispositions to classify, categorize, infer, or apply. While I am invested in questions concerning the semantic approach, such as which metasemantic theory stands best to reason, my primary interest will be in the psychological approach. In other words, I will, more or less, treat concepts as knowledge structures, systematically bound bodies of information stored in long term memory – where 'knowledge' is to be understood not as factive, but rather 'any contentful state that can be used in cognitive processes' (Machery 2009, p. 8). Specifically, I will draw attention to two aspects of such knowledge structures: *characterizations* and *classificatory dispositions*.

When it comes to social reality, the former enables individuals to intuitively engage with their environment. Such bodies of information consist of attributed features to a subject matter, often affectively-laden, and structured according to which features stick out relative to others, and which features play a stronger explanatory role. For example, my characterization of a schoolyard bully is

[8] Perhaps this is not so obvious. Herman Cappelen (2018) argues against this, to a certain extent. Conceptual engineering, on his account, is not about changes to concepts *per se* but to the intensions and extensions of terms.

one that consists of the features of being mean and brutish, and such features explain why bullies are feared. What is important to note is that characterizations provide information about the significance of a particular social kind or category.

The latter involves our mental and linguistic dispositions to treat or see aspects of the social world as falling under the same kind or category (e.g., who we classify as a 'bully'). And, importantly, characterizations and classificatory dispositions are in complex interaction with one another. That is, what we take to be significant about a social kind or category plays a substantive role in shaping our beliefs, opinions, and behaviour with respect to which aspects of social reality belong to the same kind or category.

The reason for thinking of concepts in this way relates to what it is that conceptual engineering is supposed to achieve. As someone who is interested in the amelioration of unjust social structures through the revision of our tools for thought and talk, concepts, understood as bodies of information and cognitive dispositions, provide a location at which this is possible. In other words, how we classify aspects of the world makes a difference to the organization of social behaviour. Put differently: better classificatory practices will be apt to bring about better social relations. I will save discussion of this for the next chapter.

Before I continue, I want to address a potential problem. Some, like Steffen Koch (forthcoming), will object that focusing primarily on psychological phenomena *alone* will not suffice to meet the relevant aims or goals of conceptual engineering, whatever they might be. This is because if we focus solely on changing non-semantic mental facts, say facts about classificatory dispositions, this will bring about 'linguistic confusion' insofar it will 'lead people to use words incorrectly and thus to assert many falsehoods' (ibid. forthcoming a, p. 5).[9] First, for one who is less interested in semantic facts, the problem of 'incorrectly' using words and 'asserting falsehoods' is plausibly not very interesting. It seems perfectly reasonable for someone who *simply* cares about mental and linguistic behaviour,[10] independent of semantic facts, to be rather indifferent to the correct use of words or the truth of expressions. More so, one might think that while semantically incorrect, insofar as the use of the words does not correspond to, or align with, its actual public meanings or semantic values, the new use is morally, socially, or politically correct – and meaning must catch up.[11] But withstanding this, Koch's objection only holds if there is not *enough* change to the linguistic behaviour of speakers within a community. That is, linguistic *confusion*, understood as a thoroughgoing psychological phenomenon, will occur only when there is a lack of robust uniformity in the use of a term or expression within a community. But one might wonder if there are any conceptual engineers who just want to change *some* people's usage of a term or expression?[12] This seems highly unlikely. The goal of conceptual engineers, it seems to me, is to change *all* or *most* people's usage within a community; especially with the aim of social justice in mind.[13] And, importantly, if this is achieved, then there

[9] This was first introduced by Jennifer Saul (2006) in her response to Sally Haslanger's (2000) ameliorative definition of gender.

[10] This is, perhaps, too narrow. Other aspects of psychological understanding may also be important, even when it is not fully manifest in linguistic *behaviour*. After all, we do care about what other people think, whether or not this translates into action.

[11] Arguably, meaning will change as our dispositions do. Though, there will perhaps be a time lag.

[12] In saying this, some might be interested in changing the dispositions of dominant speakers to align with the concepts developed from marginalized situatedness. In this case, however, it is not that the aim is only to change some part of the linguistic community such that one part speaks one way and the other speaks another way. In other words, the goal is uniformity.

[13] Of course, for some theoretical concepts, even for the purposes of social justice, we might only intend to affect *expert usage*.

should be no threat of linguistic confusion. After all, usage of a term or expression will be robustly uniform, whether or not it is 'correct' or expresses 'truth.'

Koch will likely respond that the relevant confusion is not one that is 'psychological' in the sense that there is a conscious lack of understanding about how to use a term or expression, but rather that one loses knowledge about what they are talking about – it is entirely possible that, through semantic change, the subject matter of a concept can shift without our knowing.[14] I think the use of 'confusion' here is misleading. Sure, there might be a mismatch between what a speaker believes they are talking about and what they are in fact talking about. However, to reiterate, this surely isn't a problem for someone who is *just* interested in changes to the mental and linguistic behaviour of a community of thinkers and speakers – there is no confusion about usage, just mistaken beliefs or opinions about truth and correctness.

With this on the table, I believe it is no problem for conceptual engineers to focus squarely on the revision of non-semantic mental and linguistic facts.[15] Of course, a failure to achieve the goal of changing the mental and linguistic behaviour of all or most members of a community will be apt to cause linguistic confusion in the psychological sense. However, this is not a failure to identify the right philosophical methodology of conceptual engineering, but rather a failure for one to complete their ameliorative project. We should be careful to tease these apart.

### 1.1. Conceptual Engineering

After forming a belief, we may come to find that it stands in need of improvement. In what sense? A simple answer is that it doesn't accurately represent how things are in the world. For example, you might have thought that whales were fish until you took a class in biology and discovered that they are mammals. In the face of a defective belief, what *should* one do? This normative question appears to have an intuitive answer: One should revise their beliefs to correspond with the facts.

Like belief, our more basic ways of representing things can stand in need of improvement. Kevin Scharp (2013), for example, has argued that the ordinary concept of truth cannot be assigned a consistent extension. This, in turn, leads to certain paradoxes. Thus, we are faced with a normative question: In the face of a defective concept, what should one do? According to Scharp, one should, in certain contexts, replace the ordinary concept of truth with two proposed alternatives – this will help us to avoid running into theoretical roadblocks, enabling advancements in our understanding of key philosophical issues. The goal of revision, for Scharp, is *epistemic*. It is a matter of ensuring that we assign a consistent extension to our concept(s) of truth.[16]

Concepts can stand in need of improvement in a variety of other ways. Sally Haslanger (1993, 2000) has argued that we should revise the concept of woman on the grounds that it would serve the interests of feminist inquiry, therefore advancing pursuits in the project of social justice. The goal of revision is social, moral, and political. Projects in a similar spirit include: Robin Dembroff's amelioration of SEXUAL ORIENTATION, Kate Manne's revisionary analysis of MISOGYNY, Katharine Jenkins's trans inclusive account of WOMAN. Other examples abound. It is such projects, namely those that concern *social kind concepts*, that will be the focus of this thesis. Social concepts, roughly speaking, describe and regulate aspects of social life. Paradigm examples include: WOMAN, MAN,

[14] Cf. Sarah Sawyer (2020).

[15] Others have made a similar point, such as Isaac (2020) and Fischer (2020).

[16] For similar projects, see Clark and Chalmers (1998) on *belief*, and Machery (2009) on *concept*.

NON-BINARY, BLACK, WHITE, MIXED RACE, FOOD, GAY, SEXUAL HARASSMENT, GASLIGHTING, etc. How such concepts are able to play this role in social life will become clearer in the next chapter.

Evaluating concepts, relative to the aims of inquiry (e.g., epistemic, moral, political), is often called *conceptual ethics.* And, after a concept is found wanting, the next step is to propose how a concept can be improved, and to offer suggestions for how to implement such improvements. This is called *conceptual engineering.* For ease of expression, I will follow the trend of calling this whole process, 'conceptual engineering.'[17]

For the most part, this chapter, and the thesis more broadly, is an exercise in the *metaphilosophy of conceptual engineering*. That is, I will not so much engage in conceptual engineering *qua* conceptual removal, revision, replacement, or introduction. Rather, I will discuss two ways of thinking about conceptual engineering – *what it is about* and *what it is supposed to achieve*. Specifically, my concern is whether conceptual engineering is a plausible means achieving, or aiding in the process of, social justice. I believe that it is. But we must be careful to understand how and why. To this end, I will offer my preferred account of conceptual engineering, and I will discuss its role in social justice theorizing.

## 1.2. What is Conceptual Engineering About?

It should come as no surprise that our existing set of concepts could be improved. Recent history of philosophy is replete with examples of theorists offering alternative ways of better understanding a subject matter. To reiterate, normative projects of this kind fall under the banner 'conceptual engineering.'

But, when one claims that 'our existing set of concepts could be improved', what does one mean? There are a few things to ask: (1) What is a concept? (2) Who is 'our'? and (3) What kind of improvement? To get (2) out of the way, I will simply state the domain of conceptual engineering that interests me. I am concerned with the conceptual repertoire of communities of thinker and speakers, rather than individuals or particular local groups. In other words, concepts, as I will understand them, emerge within a community – especially, *social* concepts. This includes whole societies such as Australia or the U.S. But, throughout the thesis, I will explore how new concepts are created and developed *within* local communities, and how it is that such concepts can take root and propagate in dominant contexts. In particular, I will address this in chapter three. For now, I will discuss (1).

The question asked by (1) represents the subject matter of conceptual engineering – what it is *about*. And, to put it bluntly, it is about concepts. However, 'concept' is a notoriously ambiguous and contested notion. Broadly speaking, theories of concepts divide into two camps: those that understand concepts as the public linguistic meanings (intensions and extensions) of non-indexical terms[18] and those that understand concepts as psychological bodies of information or knowledge structures.[19] Call the former *the semantic approach,* the latter *the psychological approach.*

---

[17] Note that conceptual engineering needn't follow this process step by step. Sometimes conceptual engineers will simply want to introduce a new concept, without evaluating an old one, perhaps because they are trying to fill in a lacuna in our existing conceptual repertoire. Moreover, Burgess and Plunkett (2020) have argued that conceptual engineering is a process that *includes* conceptual ethics, as well as conceptual innovation and implementation.

[18] Here I am restricting my focus to concepts with a stable reference. However, I acknowledge that it is common to accept that there are indexical concepts.

[19] It's important to note that not everyone thinks that concepts are the meaning of non-indexical terms. For example, Sarah Sawyer believes that the intension of concepts, which is not meaning, is determined by non-

With respect to the semantic approach, I am rather neutral as to what meanings are, or how such meanings are determined.[20] This isn't really important for what's to come. Part of the reason for avoiding talk about meaning, or the semantic approach more generally, is that a lot of the discussion in recent conceptual engineering literature has been bogged down in questions concerning metasemantics (e.g., Cappelen 2018; Haslanger 2020; Pollock 2020; Thomasson 2020; Ball 2020; Sawyer 2020; Deustch 2020; and more).[21] Such questions, I take it, *are* important to the extent that, if one's goal is to change meaning, then it matters what theory of meaning is correct. After all, we might find out that meaning change is simply too hard (Cappelen 2018; Deustch 2020), or reasonably easy (Pinder forthcoming), or that it requires collective long-term control (Koch forthcoming b), or that it depends on future use (Ball 2020), etc.

As stated, my interest isn't in meaning understood as stable semantic contents of public language expressions. And it isn't clear to me what the standards are for determining which theory of meaning is correct. With this said, I acknowledge that meaning change is something that is often a matter of moral, social, and political concern. For example, I take it at face value that when marriage equality was being promoted in Australia, advocates *really were* aiming to change the meaning of 'marriage.' However, I suspect that if such advocates were to be told that meaning change is beyond control, even collectively, that they would settle for simply changing *actual mental, linguistic, and social behaviour* whether or not this translated to a change in definition, intension, extension, reference, or what have you.

One might stop me at this point and arguing the following: there is no sense in which there can be actual change to the mental and linguistic dispositions of a community without a change in meaning. This might be because, say, meaning *just is* determined by patterns of prior and actual usage within a community; or because meaning is a function of certain internal states, such as inferential dispositions that ground linguistic behaviour. If this is the case, then so be it. Meaning change is just a happy consequence of the kind of change that interests me in this thesis. However, all that I need to accept is that *if* mental and linguistic dispositions with a term can deviate from the meaning of that term, then my concern with the former, not the latter.

There are a number of theorists who have discussed the idea that meaning can come apart from mental and linguistic dispositions; or what Steffen Koch calls *Divergence* (forthcoming a).[22] Most prominently are those who endorse versions of semantic externalism. For example, consider Kripke's (1970) causal theory of reference. This tells us that reference is fixed through an initial dubbing, a baptism, with a proper name or term to an individual or natural kind. When fixed, reference is then sustained through causal chains of communication with the relevant name or term that is 'hooked up' with the world, as it were. The upshot is that the knowledge, beliefs, or other attitudes a speaker associates with a name or term play no role in the determination or continuation of reference. That is,

---

descriptive relations to objective properties in the world; whereas word meanings are determined by patterns of usage and deference to competent speakers. So, while words *express* concepts, concepts are *not* the meanings of words. And, I should say that I find this convincing to a certain extent – moreover, it is completely consistent with what I am arguing for. After all, when I say that I am interested in mental phenomena, such as classificatory dispositions, I don't need to say that such things *are* concepts.

[20] Though, I accept that meaning determines reference.

[21] Cf. Mari Mikkola (2009), who argues that, in response to Haslanger, we needn't bother trying to provide ameliorative definitions of gender terms or concepts – we can avoid semantics altogether. Instead, we can rely on 'extensional intuitions'; intuitions about which individuals we think a term applies to. I take my project to be somewhat in the spirit of Mikkola's suggestion.

[22] Jared Riggs (2019) is a prominent recent example of this line of thinking.

speakers can be mistaken about to what a name or term refers and yet still refer with that name or term.

Similarly, Putnam (1975) offers a form of externalism that allows for ignorance about reference and meaning. On Putnam's way of thinking, the environment in which one is situated plays a substantive role in determining the reference of a term. This means that two people in the same mental states can refer to different things depending on contingent environmental factors. Burge (1979) extends this thought to the social domain, arguing that experts and lay people within the same linguistic community can refer to the same thing even with drastically different states of mind.[23]

So, what's the take away from this? In a nutshell, the main conclusion is that meaning does not amount to understanding. And this is not controversial in the existing literature. Nevertheless, I want to acknowledge that metasemantics must have *some* grounding in actual understanding or dispositions. Put differently, actual understanding, or mental and linguistic dispositions with a term or concept, *influences* meaning – the question is just, 'how much?' Understanding is a first step for meaning change, no matter one's preferred theory.

But even with this in mind, let me reiterate that, for the most part, I will not be interested in meaning. The natural next question is: Why care about mental and linguistic dispositions? Briefly, such dispositions have substantive practical and epistemic effects on social life. Many of us care about the real world consequences of our engineering projects. I'll put this question on the backburner, and I will address it at the end of the next chapter when I discuss the different reasons for undertaking conceptual engineering projects – this will provide an answer to (3) above. For now, I will spent time explaining the kind of mental and linguistic dispositions that will be the focus of this thesis.

If not meanings, what's are concepts? Concepts have been also understood to be psychological bodies of information, or knowledge structures. Theories that construe concepts in this way I have called 'the psychological approach' – an approach that is characteristic of the study of concepts within the field of psychology. Such psychological phenomena, to some extent, have not been the focus of much philosophy. Of course, some philosophers are interested in the relationship between psychology and the nature of concepts. But the primary focus in philosophy has been on the role of concepts in securing reference and determining truth conditions. So, what makes the psychological approach distinct from philosophical theorizing about concepts?

Psychologists who are interested in the nature of concepts tend to focus on the relationship between bodies of information and cognitive competencies. The primary goal is to 'determine what kind of knowledge is used by default' in certain mental processes (Machery 2009, p. 34), where, to reiterate, knowledge is understood not as factive.

Such processes are not those that concern, say, perception or motor skills. Instead, the relevant competencies are *higher order*. This is the aspect of cognition that tends to be 'under intentional control; their products are often (or can be) conscious; and they tend to be slower than the processes underlying [perception and motor skills]' (ibid. 2009, p. 9). This includes classification or categorization, inference making, analogizing, planning, reasoning, metaphorical perspective-taking, etc.

[23] And *Divergence* is also possible on internalist accounts of meaning or reference (Koch forthcoming a). This is because, *inter alia,* it is entirely plausible that the cognitive states that determine reference, or at least some of them, are inscrutable or not entirely consciously accessible (Cappelen 2018, p. 82).

Higher order cognitive competences involve internal or private processes of accessing information that is stored in long-term memory[24] in order to fulfil the functions that are characteristic of such competences. Lawrence Barsalou, a prominent psychologist of concepts, captures this thought succinctly:

> ... a concept, roughly speaking, is knowledge about a particular category (e.g. birds, eating, happiness). Thus knowledge about birds represent the bodies, behaviours and origins of the respective entities. *Knowledge plays a central role throughout the spectrum of cognitive activities. In on-line processing of the environment, knowledge guides perception, categorization, and inference.* In off-line processing of non-present situations, knowledge reconstructs memories, underlies the meanings of linguistic expressions, and provides the representations manipulated in thought (Barsalou et. al. 2003, p. 84 my emphasis).[25]

Taking stock of such ideas in the psychological theorizing of concepts, Edouard Machery argues that the theoretical term 'concept' is often used as follows:

> A concept of $x$ is a body of knowledge about $x$ that is stored in long-term memory and that is used by default in the processes underlying most, if not all, higher cognitive competences (2009, p. 29).[26]

The details of Machery's interpretation of how 'concept' is often understood in psychology aren't important to our discussion. What *is* important is the idea that we draw on bodies of information, or knowledge structures, that are stored in long-term memory and which guides, or plays a substantive role in shaping, online higher order cognitive processing. To continue with the example above, our knowledge of birds, or the psychological information to which we have access, such as information about their bodies, behaviour, and origins, is used in online cognitive processing in understanding and interaction with birds.

Knowledge structures, or systematically bound bodies of information, will be part of the working definition of concept in this thesis. Additionally, I take the mental and linguistic dispositions that such knowledge structures underlie, or explain, to be a part of the notion of concept as well. In particular, I will focus on two aspects of our psychology: *characterizations* and *classificatory dispositions.* Briefly, the former are *structured* bodies of information that partly constitute our understanding of the significance of a particular kind or category. I will have much more to say about characterizations in the next section. For now, what I want to discuss is classification. This is a higher order cognitive competence consisting of certain (implicit) beliefs or opinions about the things that fall under the same kind; or our disposition to treat or see the aspects of the world as belonging to the same category. To see this more clearly, let's explore an example. Classificatory dispositions involve the recognition of, say, whether a particular guitar is a Gibson. Further, such dispositions involve the use of words under appropriate conditions, such as calling something a 'Gibson' when in the presence of someone who wants to know what kind of guitar one plays. And, even further, our

[24] *Affective* processing, which is an important aspect of characterizations as will see later in the chapter, is perhaps not always accessed in long-term memory (Le Doux and Brown 2017).

[25] Similar statements are made by Solomon, Medin, and Lynch (1999, p. 99), Komatsu (1992, p. 500), Smith and Medin (1981, Ch. 1), Barsalou (1989, p. 76), and others.

[26] Machery explores different psychological accounts of concepts, such as proto-type theories and theory-theories.

classificatory dispositions involve certain inclinations to draw inferences on the basis of (implicit) beliefs or opinions about a kind or category – such as inferring that Gibson guitars are instruments, or that they are expensive.[27] Such dispositions, as well as the bodies of information, or knowledge structures, that underlie such dispositions, are a part of the notion of concept that I am working with in this thesis.

A possible problem with this idea is that Machery's definition of concept tells us that bodies of information *underlie* cognitive competences. This seems to suggest that particular cognitive processes, such as classification or categorization, only follow from, or are guided by, concepts. In other words, such cognitive processes are not *a part* of concepts understood in the psychological sense. The idea, it seems, is that in order to get stable dispositions, such dispositions must be grounded in some underlying cognitive structure that explains that stability. So, stable classificatory dispositions are one thing, and the knowledge structures that underlie such dispositions are another.

However, despite being different elements or aspects of psychology, there appears to be a rather intimate relationship between bodies of information and our higher order cognitive competences: stability of disposition *requires* stored knowledge structures. Given this tight-knit relationship, I will simply treat both bodies of information *and* the cognitive dispositions that they underlie as falling under the notion of concept; at least for this thesis. To reiterate, I will be primary interested in concepts understood as consisting of characterizations and classificatory dispositions.[28]

This raises a pressing question: How are concepts, understood in this sense, individuated? I don't wish to spend too much time on this, so I'll only briefly explore a couple of options. One answer, a broadly Fregean way of thinking, might be that individuation is a matter of sameness of knowledge structure – however fine-grained the matching relation must be. What does this mean? Of course, no two individuals will associate *precisely* the same knowledge structures with a word; this is because they will, through their unique experiences, have idiosyncratic beliefs, attitudes, and classificatory dispositions. However, the Fregean approach requires that there be some *core* aspect of their knowledge structures that constitutes their competence with a concept. And to share the same concept, the story goes, is to have precisely matching core knowledge structures.

The problem is that this way of thinking, the Fregean story, doesn't appear to be promising for my purposes. I want to individuate shared concepts in a way that allows for significant variability in knowledge structure within a community. Given this, I will commit to a theory of individuation that takes psychological concepts to be individuated via epistemic or communicative relations. That is, taking inspiration from Samuel Cumming (2013), concepts are individuated via *de facto* or automatic coordination in which internal knowledge structures 'are associated by communicative "conventions" within a group of agents" (ibid. 2013, p. 7) – policies of communication in the service of a shared goal (e.g., referential coordination).[29] This does not require that the knowledge structures that different individuals possess within a community precisely match up in some core aspect, but rather that such individuals sit in the right communicative and epistemic relations, a representational

[27] In an unpublished manuscript, Jennifer Nado argues that such dispositions are among the primary concerns of conceptual engineering interventions – in particular, social justice projects such as the concepts surrounding marriage equality.

[28] For one unconvinced by this, another response is that while the notion of concept that I have introduced might go against the typical understanding of concept in psychology, it seems available to me to simply engineer a new notion of concept to help illuminate what the projects of conceptual engineering are, or should be.

[29] This kind of approach has come to be known as the (mental) file-network theory. An early proponent is John Perry (1980, 2001); and, more recently, Laura Schroeter (2012). Such accounts are intended to contrast Fregean matching theories with relational views.

tradition, via shared natural language expressions. In a little more detail, the idea is that individuals are committed to stably interpreting a particular word with one of their internal knowledge structures, and the result is that different individuals' knowledge structures are stably wired together in virtue of being related to the same public language word.

Getting into the complicated details of this picture, such as the specifics of how *de facto* coordination is achieved, isn't needed. Suffice it to say that the general picture is that concepts, understood as knowledge structures that include bodies of information and classificatory dispositions, are shared over time within a community, and coordination of these structures requires natural language such that associations develop within this community which align patterns of understandings with the same public term or expression.[30] To separate coordinated networks of concepts, or those concepts that are *de facto* coordinated via natural language signalling strategies between agents within a community, from the idiosyncratic concepts that are not coordinated, call the former *coordinated concepts.* It is such concepts, coordinated bodies of information and classificatory dispositions, or systematically linked networks of individual knowledge structures that underlie classification, that will be the primary focus of this thesis. What's important to note is that coordinated concepts accommodate for a wide range of variation in knowledge structure within a community while allowing that individuals are talking about the same thing. Despite this, communication and common environment will tend to propagate *overlapping* patterns in the knowledge structures within this community.[31]

Importantly, this notion of concept is what I will take to be the subject matter of conceptual engineering. This is not to say that *all* conceptual engineering projects are, or must be, about revising particular bodies of information and classificatory dispositions. This is, of course, not true. As said, some, plausibly most, conceptual engineering projects are preoccupied with conceptual change *qua* meaning change (e.g., Haslanger 2000, 2020; Cappelen 2018; Thomasson 2020; Sawyer 2020; Richard 2020; Ball 2020; Pinder forthcoming; Koch forthcoming a; forthcoming b, etc.). Roughly, projects of this kind have the following form:

(1) Conceptual Ethics:

(i) Identify a concept *C* to be evaluated, and discern its meaning $M_1$.

(ii) Make a judgment about whether *C* should be revised given an assessment of $M_1$ relative to the aims of the practice.

(2) Conceptual Engineering:

[30] Compare this view with another alternative. Vincente and Martinez (2016) argue for a different psychological theory of concepts that is friendly to hybrid multi-stranded concepts. They individuate concepts in terms of *functional stable coactivation* of representations. The idea is that representations belong to the same concepts when: (i) they are activated concurrently by the same pattern of stimulus; (ii) what is activated has some functional significance for the task at hand, and (iii) the coactivation patterns remain substantially stable in tasks related to the same category (Vincente and Martinez 2016, p. 67 – 68).

[31] However, if someone's knowledge structure is completely out of sync with the rest of the community, the coordination relation will break down. For example, suggested to me by Laura Schroeter, is Mrs. Malaprop who thinks that allegories live on the bank of the Nile. Once we find out about how she understands 'allegory,' we abandon the policy of coordinating with her – we no longer interpret her homophonically.

(iii) Propose that C should have an alternative meaning $M_x$, where $M_x = M_2, M_3, ..., M_n$ depending on the meaning that is judged to be best relative to the aims of the practice.

(iv) Suggest ways that C could come to mean $M_x$.

We might call this *The Standard View.* The goal is to take a look at the meaning of a term or concept, judge whether it stands in need of improvement, and if one decides that it does, then a proposal for an alternative meaning is offered. It should be noted that *The Standard View* is a rough schema that does not capture all projects of meaning revision. Some might want to introduce a new concept into an existing repertoire; some might want to eliminate a concept. The point that I want to make explicit is that, on *The Standard View,* meaning is the name of the game.

It seems to me that to think that conceptual engineering is *just* about meaning (i.e., intensions and extensions) is a critical oversight of important features of our thinking and speaking practices. Put differently, *The Standard View* is far too parochial. It is short-sighted about the variety of ways that terms or concepts are involved in complex forms of thought, communication, and coordination. Thus, in order to accommodate such complexities, I will take the project of conceptual engineering to consist of the following schema:

(1*) Conceptual Ethics:

(i) Identify a (coordinated) concept C and discern its existing pattern of understanding and use $U_1$.

(ii) Make a judgment about whether C should conform to $U_1$ relative to the aims of the practice.

(2*) Conceptual Engineering:

(iii) Propose that C should have an alternative pattern of understanding and use $U_x$ where $U_x = U_2, U_3, ..., U_n$ depending on the understanding and use that is judged to be best relative to the aims of the practice.

(iv) Suggest ways that C could come to be understood and used in way $U_x$.

We might call this *The Coordinated Psychology View.* On such a view, conceptual engineering is about changing coordinated patterns of understanding and mental/linguistic dispositions within a given community. To be clear, there are all sorts of bodies of information and dispositions that might

be the focus of conceptual engineering.[32] As I have said, my interest is specifically in characterizations and classification.[33]

At this point, I want to respond to some possible objections. The first is this: One might say that conceptual engineering *really is* just about meaning change, and that anything else is another kind of project – i.e., cognitive engineering, epistemic engineering, social engineering, classification engineering, etc. If this is true, then that's fine by me. However, it would strike me as odd if the label 'conceptual engineering' laid claim only to meaning and reference given its historical roots in social justice theorizing. After all, the goal of Haslanger's (2000) revisions of race and gender terms doesn't stop at meaning or reference change – and, plausibly, if meaning and dispositions come apart, it needn't require such change at all. The goal is to have an ameliorative impact on social facts; real world change to our social behaviour. So while we might accommodate meaning and reference change as a goal of conceptual engineering, it isn't the be-all and end-all. And meaning change needn't even be thought of as central to conceptual engineering projects. Conceptual engineering includes, more broadly, our use of terms or concepts, and the understanding that underlies such dispositions. Moreover, it is through changing such patterns of understanding and use that social change can be achieved. I will say more about this in the next chapter.

The second objection is this: I should not dismiss the importance of reference when thinking about mental and linguistic dispositions. This is because such dispositions are grounded in bodies of information, or knowledge structures, that are *about phenomena in the world*. That is, a concept must have an extension that our thought and talk is about – even if the extension is empty, or we're mistaken about what is in the extension. Part of the reason for this relates to something that was said at the beginning of this chapter: we might be using words incorrectly or not expressing truths (Koch forthcoming a). So, without reference it will be hard to normatively assess, on epistemic grounds, our mental and linguistic behaviour.

To reiterate, for someone who is less interested in truth, this might not be such a problem. Truth needn't be the normative yardstick for the appropriate use of a term or concept. Sometimes we might simply want to assess the appropriate usage of a term or concept relative to social, moral, or political standards – standards that needn't align with the 'true' meaning or semantic content of the term or concept.[34] Nevertheless, I accept that concepts must have semantic structure, that which fixes the reference of a concept. And, moreover, concepts must have psychological structure, that which plays an explanatory role in how we classify.[35]

Even with this said, I will be mostly indifferent to the semantic structure of concepts. I accept that concepts must refer, but I leave it more or less open as to how concepts get their referents. My focus

[32] Such as traditional views that take concepts to mentally represent necessary and sufficient conditions; proto-type theories that take concepts to mentally represent clusters of typical features; atomistic views that concepts to be symbols in the language of thought; and theory-theories that take concepts to mentally represent theories.

[33] And other projects will be interested in psychological dispositions beyond classification (e.g., Riggs 2019, Pollock 2020, Isaac 2020, Fischer 2020, etc.).

[34] Moreover, as Podosky (2018) argues, conceptual engineering can be about *creating social facts*, and therefore making certain statements true.

[35] This could be construed as getting us close to a *dual-content* account of concepts (Margolis and Laurence 1999; Weiskopf 2009; Del Pinal 2015; Camp 2015). In other words, conceptual content comes in two forms: semantic and psychological. And both play important roles in thought and talk – semantic content determines whether we are thinking and speaking in accordance with truth; psychological content undergirds our mental and linguistic competencies. Despite this, I leave it open how to best construe the relationship between semantic content and psychological content.

will be primarily in the psychological structure of concepts; the bodies of information that underlies classificatory dispositions, and the like. It is worth mentioning, however, that the semantic content and psychological content of concepts are only *contingently* connected (Margolis and Laurence 1999, p. 42). Importantly, our psychological associations with a subject matter, and our dispositions to classify things as falling under the same kind or category, needn't contribute to determining reference. So, following Elisabeth Camp (2015), I believe that a useful metaphor for thinking about the relationship between such contents is the following: the psychological content of a concept is the 'roiling electron clouds' orbiting more stable nuclear semantic structures. That is, the psychological content associated with the same subject matter can vary between members within a community, while the semantic content remains the same.[36]

Despite this variance within a community, there are roughly *matching* knowledge structures – where this 'matching' does not contribute to the individuation conditions of concepts. To foreshadow what's to come, similarity of coordinated concepts can be regulated or stabilized via environmental mechanisms of psychological convergence. So, on my way of thinking, psychological concepts are shared not just in the sense that people enter into *de facto* coordination via signalling strategies involving natural language expressions. Rather, similarity of characterizations and classificatory dispositions emerge within a community of thinkers and speakers.

### 1.3. Characterizations

So far I have said that 'bodies of information underlie classificatory dispositions.' This, as it stands, is pretty vague. It's important to recognise that the bodies of information, or knowledge structures, that guide our classificatory dispositions are not simply amorphous and undifferentiated. Rather, they are *structured*. And bringing attention to this structure will enable us to identify distinctive locations for conceptual engineering. Importantly, if we change the psychological content used in higher order cognitive processing, then we can bring about changes to classificatory dispositions – at least with respect to social kinds or categories. But which content, or which aspect of our conceptual understanding, should be revised will depend on where that content figures in our overall thinking practices.

To discuss the structured bodies of information that play an explanatory role in online cognitive processing, I will draw on the work of Elisabeth Camp (2015). She calls such information structures *characterizations*. I will follow her in using this terminology.

The fundamental role of a characterization is to enable a thinker to engage intuitively with the cognitive and environmental context in which she is embedded. To perform this role, characterizations have three main features.[37] The first is their *content*. Characterizations are rich collections of attributed features or properties to a subject, or set of subjects, where such features or properties are often affectively laden (ibid. 2015, p. 602).[38] For Camp, this content can range anywhere between general and specific. For instance, think of a Melbourne hipster. My

[36] There is disagreement over this. For example, Hilary Putnam (1975) has argued that psychological content such as stereotypes contribute to meaning in virtue of playing an essential role in our mental and linguistic behaviour.

[37] There are competing accounts of information structures that are like characterizations. For example, there are *classical theories*, *prototype theories* (e.g., Rosch and Mervis 1975; Rosch 1978; Hampton 2006), and *theory theories* (e.g., Carey 1985, 2009; Gopnik and Meltzoff 1997; Keil 1989). I will focus solely on characterizations.

[38] Camp also discusses the notion of *characterization* elsewhere. See: Camp 2013, 2017a, and 2017b.

characterization of a hipster includes being aloof, environmentally conscious, and marginally snobbish. More specifically, hipsters have beards, are overly concerned with coffee, and dress in a way that resembles 90s fashion. Further, Camp stresses the importance of affectively laden properties concerning how the subject in question tends to make one feel. A hipster might make one feel sheepish about one's fashion choices or excitable character.[39]

The second main feature is *endorsement*. It is not a requirement of characterizations that their subjects actually possess the properties ascribed to them (ibid. 2015, p. 604). Taking our example of a hipster, one can have an excitable character, or not be so interested in blends of coffee, and still be a hipster. Such properties are only considered to be *fitting* for hipsters insofar as they *should* have them.

The third and last feature of characterizations is their *structure*. This involves at least two dimensions of psychological importance.

One dimension is that the attributed features or properties of a characterization are more *prominent* than others. This is roughly in line with Tversky's (1977) notion of salience, consisting of intensity and diagnosticity. An attributed feature is intense insofar as it has a 'high signal-to-noise ratio', that is, the feature sticks out relative to a background of other features (Camp 2015, p, 605). And an attributed feature is diagnostic insofar as it is a useful means of classifying things into certain categories (ibid. 2015, p. 605). What's important to note is that both intensity and diagnosticity are highly context-sensitive. There are contexts in which typically prominent features are unusually common and hence will not stand out relative to other features, and, in virtue of this, such features will not serve usefully in our classificatory practices.[40]

The other dimension of psychological importance is that some features are more *central* than others. The person attributing features to an object treats such features 'as causing, motivating, or otherwise explaining many of the subject's other features'[41] (ibid. 2015, p. 605). Think back to our example. That a hipster is aloof might explain why he is admired or found attractive.[42] For Camp, we can tell how central a feature is by counterfactually reflecting on how much else about the subject one thinks would change if that feature were taken away (ibid. 2015, p. 605).

Taking these two elements together, structures of prominence and centrality are highly intuitive and holistic. The role a feature plays depends on the role of other features. To see this, we need only think about the role of characterizations in forming Gestalt ideas. Characterizations help in framing and organizing one's thinking about a subject matter. This includes what one notices, one's affective responses, and one's understanding of relevant action scripts, etc. For example, if we take the (affectively-laden) property of being aloof as playing a structural role in our representation of a hipster, it will in turn imbue other properties with different emotive and evaluative significance. We might reason that someone is quiet because they are aloof, instead of being socially anxious. We may further judge this behaviour to be snobbish. In this case, the features that I attribute to the hipster are

[39] Camp maintains that characterizations needn't be about social types (i.e. stereotypes), but can represent individuals. I am really only interested in the former. Though, perhaps a focus on individuals might be a means of imagining possible ways that the world could be that would serve the interests of social justice theorists.

[40] For example, that S is a hipster at a party full of hipsters means that S being aloof is not going to stick out relative to S's other properties (namely, those properties *not* characteristic of hipsters). Moreover, if someone who is looking for S wants to locate them amongst hipsters, then they won't be (only) looking for the person who is aloof.

[41] See Thagard (1989), Sloman et al (1998), Murphy and Medin (1985).

[42] At least within the hipster universe.

themselves a holistic function of the roles that these features play in relation to the rest of my characterization.[43]

As we can see, the psychological content used in online cognitive processing often has a particular and unique structure. And how such content is structured makes a difference to our classificatory dispositions. In particular, if a feature is prominent, then this will motivate us to classify in one way rather than another. For instance, in a room of people with mostly black hair, the property of having red hair will be prominent. And owing to its prominence, classifying people on the basis having red hair will be a useful means of picking specific people out from a crowd (e.g., 'she's the one with red hair'). However, in a room of people with red hair, the property of having red hair will not be prominent. Classifying on the basis of having red hair will be less useful.

This example gets to the heart of the relationship between characterizations and classificatory dispositions to which I want to draw specific attention. Characterizations are structured bodies of information about a subject matter, or dispositions to construct a way of seeing at a time, that enables us to grasp what's significant about a category or kind (e.g., being able to reliably pick someone out from a crowd). This puts characterizations in complex interaction with our classificatory dispositions. What we take to be significant about a kind or category will play a substantive role in underlying how we classify aspects of the world – the social world, in particular. For example, if we think that what's important about the social kind *woman* is a particular self-identification or self-conception, then this will make a difference to whom it is we classify *as* women – importantly, it will include trans* women (and exclude trans* men).[44]

With this said, however, it's important to be clear that 'fittingness' and typicality effects show that characterizations don't *directly* determine classificatory dispositions. We have certain theoretical commitments, and other bodies of information, that influence our classificatory dispositions that can come apart from our characterizations or stereotypes of a particular subject matter. For example, despite our characterization of fish being such that it fits with the features we might be inclined to attribute to whales, our commitment to trusting scientific experts means that we tend classify whales as mammals (and not fish). However, with respect to social kinds, there seems to be a much more intimate relationship between characterizations and classification. That is, our characterizations of a social kind or category are at least *apt* to influence or guide our dispositions to classify aspects of the social world as falling under the same kind. Put differently, what's important to our understanding of a subject matter tends to make a difference as to whether we treat or see a collection of things as belonging to the same category (e.g., emphasising the significance of self-identification with respect to the category *woman*).

How does this relate to conceptual engineering of social kind terms? If we are to change the classificatory dispositions of a community of thinkers and speakers, there is a sense in which we will

---

[43] To be clear, I am understanding characterizations not to be abstract entities, but rather *psychological* dispositions, or the products of such dispositions, to see things a certain way. So what matters to a characterization is that one *actually* structures their thinking along certain dimensions of prominence and centrality. Further, there is a sense in which characterizations are *stable* insofar as we tend to have 'ramified' stereotypes that apply consistently across contexts – given certain experiences, we might be scared of dogs no matter how friendly they are behaving. But, there is also a sense in which characterizations can be changed *ad hoc* depending on which features of the context happen to stick out relative to others. Again, to reiterate, coordination of associative thought is achieved via policies of communication – this helps to stabilize our dispositions to see the world in a certain way. I will speak more about further social mechanisms that stabilize associative thought in the next section.

[44] Compare this to those who believe that what's important about the category of *woman* is having certain biological features.

need to target the shared characterizations that underlie such dispositions. It will not be enough to simply *tell* people to classify one way rather than another, but instead we must find ways of reshaping how people frame their overall thinking about a subject matter. For example, take Kate Manne's (2018) ameliorative proposal of MISOGYNY. If we want to get people to think and speak about misogyny as a structural policing mechanism rather than an attitude of bad individuals, then we must get them to change what they believe is most important about the category *misogyny* – such as assigning greater significance to certain features (e.g., how men 'correct' women's behaviour, or demand emotion attention, etc.), and ignoring or downplaying importance of others (e.g., the hatred of women). When this shift in characterization is achieved, then revised classificatory dispositions will be apt to follow. That is, we will not just rethink the significance of certain features over others, we will become disposed to think and speak on the basis of this – importantly, we will classify in accordance with the features that are prominent and central in our revised overall thinking about a subject matter (i.e., misogyny).

A lingering question remains: How do we get people to change the characterizations, or elements of a characterization, that tend to underlie their classificatory dispositions? One means by which we can achieve this is at the level of interpersonal interaction. We can alter a context so that it primes another to construct their overall thoughts, or adopt a particular point of view, with respect to a certain subject matter. For instance, if we want someone to classify something as spicy, we can bring to prominence certain things about the context such as whom the dish is being served to, and the relative spiciness of the dish compared to hotter alternatives, etc. This contextual adjustment or cognitive and affective matching is something that we do in any normal conversation; we tend to align our thinking with the overall understanding of others (Camp 2015, p. 6).

I accept that there are ways that we can bring about changes to the characterizations one is disposed to construct *in a context*. This is through controlling the features that stick out relative to others, and manipulating how such features are affectively imbued. And, this might not be such a bad means of defending oneself from pernicious ways of thinking at a given time. However, I am interested in more robust change. That is, the kind of change that targets one's characterizing dispositions *across* contexts. So, how might we achieve this? How can we change the characterizations, or elements of a characterization, that *tends* to shape classificatory dispositions of people throughout social space?

An important first step to answering this is to recognise that while characterizations are a part of one's individual thinking about a subject matter, there happens to be widespread similarity in characterizations within a community. That is, characterizations are often *shared.* For example, it is common to think of jocks as buff and brutish; and ballet as delicate and fancy – even in the face of counterexamples. So why are some characterizations, or some elements of a characterization, roughly uniform or shared within a community of thinkers and speakers? How it is that we seem to be able to draw on similar bodies of information when we engage in online cognitive processing such as classification?[45]

[45] Elisabeth Camp pushes a similar line of enquiry when she says, 'we need to further investigate forms of associative thought that are not merely idiosyncratic chains of association, but display significant structure' (2015, p. 615). However, Camp's domain of interest seems to be at the level of individuals, rather than at the level of social structures that causally exert pressure on cognitive agents to have roughly uniform associations in thought. The latter is what interests me.

## 1.4. Psychological Convergence Mechanisms

My characterization of a hipster is *my own.* It is something belongs to me, or is something that frames the way I see certain people. But what's interesting to note is that my characterization of a hipster is *not unique* to me. Instead, it seems to match the characterization that a lot of people have about hipsters; by and large, people think that hipsters really are overly concerned with coffee blends, or have beards, or wear clothes that resembles 90s fashion, etc. And, as you might suspect, this isn't just true for thinking about hipsters. It seems that there is similarity in associative thinking about other topics. For instance, many think that cats are feminine and keep to themselves; whereas dogs are masculine and uncompromisingly faithful. So, not only are we able to coordinate thought insofar as we can keep track of the same subject matter, our thoughts are also coordinated to the extent that we often have similar, overlapping characterizations; we seem to access similar knowledge structures in higher order cognitive processing. What is responsible for this?

One means of coordinating our thoughts, which we have discussed, is through policies of communication, or signalling strategies, that align patterns of understanding with the same word within a community (Cumming 2013). That is, we engage in *de facto* or automatic coordination of individuals' knowledge structures via shared and public natural language expressions. So, for example, our private associative thinking about *dogs* sits in epistemic and communicative relations with others via the shared and public word 'dog.' This, however, only tells us how coordinated concepts are *individuated* via coordination. It does not tell us how thoughts come to be roughly *overlapping* within a community – where 'overlapping' is not to be understood as providing individuation or identity conditions of concepts. Of course, words help to communicate with others, and this can contribute to stabilizing or regulating similarity of characterization within a community. For instance, through communication over time, we might come to adapt our own understanding to better conform with that of others in our community – to aid in communication and practical coordination. However, this by itself does not appear to be enough. After all, we could be completely mistaken about whether our knowledge structures are in fact similar, even with the effort to the use of the same words in order to adjust our thinking to match others over time. So, we might wonder: What else contributes to regulating or stabilizing similarity of characterization within a community?

There are different levels of explanation that would answer this question. At one level, we might be interested in something like evolutionary psychology, that coordination of associative thought is beneficial to thinkers and speakers and was thus selected-for. I don't have the relevant expertise to answer this question. [46] Perhaps this is true, though perhaps coordination of associative thought is a by-product of something else being selected-for. At another level, we might be interested in the social mechanisms that stabilize or regulate individual thinking about a subject matter – or what I will call *psychological convergence mechanisms.* And given that such mechanisms are social, the sort of mechanisms that operate within a social and representational milieu, coordination of associative thought can be achieved rather easily.

The best way to introduce the idea of a psychological convergence mechanism is with an example. Think of sex education. In a typical religious school, the lesson tends to be that sex is for reproduction; that it is between a man and a woman; and that contraception is bad. There's no

[46] However, it's worth noting that Michael Tomasselo (2008, 2014) has argued that humans, compared to other animals such as apes, are especially built for sociality insofar as we are disposed to coordinate common perspectives with the goal of fulfilling a common need. This is through capacities for social learning, imitation, matching perspectival views, recursive mind-reading, cultivating a common ground, etc.

question about what the subject matter of the lessons are: it's sex. But what students internalize from such lessons is certain information that forms the basis of their characterization, or overall understanding, of sex. Students, or rather kids, become disposed to see sex in a particular way (e.g., between a man and a woman), have beliefs about what sex is for (e.g., reproduction), and make judgments about its nature (e.g., contraception is bad). Moreover, such lessons are taught with an affective valance or colouring. Students do not just come to form the belief that same-sex intercourse is bad, their minds are shaped to *feel* that it is bad; they are conditioned into evoking feelings of disgust, aversion, shame, guilt, etc. when thinking about sex that doesn't fit the prescription (e.g., heteronormativity, sex when married, no contraception). But, here's what is true about sex: it is not just between men and women; it is not only for reproduction; and contraception isn't bad. What we can see, then, is that the mind-shaping of kids who are subject to religious education about sex will come to have overlapping characterizations through a particular psychological convergence mechanism: the religious education system.

Of course, there's no guarantee that everyone in the classroom will develop the same associations. Sometimes lessons won't stick. One might not feel persuaded by, or even listen to, the content being taught. Moreover, in time one will be presented with new information that will overwrite existing beliefs or feelings. Nevertheless, school provides a means by which associative thinking can become more or less uniform.

With this example in mind, other mechanisms of psychological convergence should be clear. The news often presents information in a way that makes repeated cognitive and affective associations between certain racial groups and crime, or laziness, or being invaders, etc. From this, people who watch and trust the news come to develop negative characterizations about such racial groups.[47] Parents often shape the minds of their kids by passing on information that they have acquired over time, and as a result kids tend to adhere to the thinking of their parents. Other mechanisms are less obvious. Some might simply have certain dispositions, rational or otherwise, and at different levels of conscious awareness, to conform to the thinking and speaking practices of the community. This might be to guarantee effective communication (e.g., talk about a subject matter without elaboration), or to ensure efficient coordination of behaviour (e.g., staying away from dog breeds that are 'dangerous'), or to reap the benefits of identifying with an in-group (Kahan et. al. 2007), etc. And there are other mechanisms beyond this.

The notion of a psychological convergence mechanism captures a range of different ways that members of a thinking and speaking community come to have similar or overlapping characterizations – beyond the use of shared words.[48] To be explicit, we can define a psychological convergence mechanism roughly as follows:

> **Psychological Convergence Mechanism:** Any social process in which there is exerted pressure for an individual to form a characterization, or an element of a characterization, about a subject

[47] Jason Stanley (2015) argues that one kind of linguistic propaganda involves repeated association between words and social meanings – such as the repeated associations, within a community, between 'Black' and criminal. Thus, when a politician says, 'We need to reduce crime in Urban neighborhoods in the U.S. context,' they are effectively smuggling in not-at-issue content about Blacks being criminals.

[48] Though, it should be noted that psychological convergence mechanisms *rely* on shared public language expressions in order to work within a community. It would be difficult for such mechanisms to operate without there being a common language that can convey the relevant information.

matter; which is apt to cause similarity of characterization with those who have been subject to the same social process.[49]

Put differently, psychological convergence mechanisms are a range of social forces that regulate or stabilize similarity of associative thought; such mechanisms reinforce patterns of thinking so that members of a community have overlapping elements in their characterizations of a subject matter. Such mechanisms are a part of our *socio-epistemic* structures – the network of epistemic relations in which we are embedded that play a role in the production and dissemination of cognitive resources, such as coordinated concepts and characterizations, meanings, standards for epistemic assessment, etc.[50] And these resources influence our beliefs, opinions, action-scripts, understandings of social possibility, and behaviour.

There are a few things more I want to say about psychological convergence mechanisms. The first is that the definition focuses on social processes that exert *pressure* on individuals. This needn't be deliberate in the sense that the aim of the mechanism is to get people to think one way rather than another. For example, news outlets are a psychological convergence mechanism, however one might think, at least once upon a time, that such outlets do not intend to shape people's minds but instead offer information for people to use in the process of reasoning to come to their own conclusion (of course, this is not true for a lot of new programmes). However, when news outlets, across the board, tend to offer the same, or roughly the same, perspective on the world then this will bring about more or less uniformity in associative thought within a community.

Second, I want to draw a rough distinction between socially *reinforced* characterizations and socially *shared* characterizations. Of course, all socially reinforced characterizations are socially shared. However, not all socially shared characterizations are socially reinforced. So, when is a characterization socially shared but not socially reinforced? It is hard to draw a sharp boundary. Nevertheless, I think that there are clear cases. If material reality is structured in a certain way, then individuals embedded in that reality will come up against similar things – and, simply in virtue of interacting or responding to the same material reality, such individuals might come to form similar or matching characterizations. For example, a society that has an abundance of carrots will likely consist of individuals whose characterization of carrots will have the feature 'being orange' as prominent and central.[51] This element of the characterization is not a product of a psychological convergence mechanism. To be clear, I am not so much interested in the characterizations that are shared in this way. My main focus, and reason for drawing this distinction, is the characterizations, or elements of a characterization, that are shared in virtue of being socially reinforced.

Third, it's important to note that this definition of psychological convergence mechanism does not say that for any two people who have been subject to the same mechanism will necessarily have the

---

[49] There are mechanisms of psychological convergence that are *not* social, such as similar sensory experiences and stability of physical environment. I am only concerned with psychological convergence that is a product of social processes.

[50] For example, Miranda Fricker (2007) discusses the idea that, owing to structural identity prejudice, our epistemic structures are set up in a way such that cognitive (or hermeneutical) resources cater mostly to dominant perspectives and experiences. Dominant group members have greater control over the means by which our collective cognitive resources are determined. Others have expressed similar thoughts. Those working on *active ignorance* often discuss the idea that unknowing is a social achievement, a matter of being embedded in epistemic relations that allow dominant group members to enact vicious epistemic behaviour in order to reproduce existing terms of engagement (e.g., Frye 1983; Mills 1997, 2007; Medina 2013; etc.).

[51] It's probably worth noting that carrots in fact come in many different colors, and that orange ones were selected-for – and the reason why they were selected for is said to be deeply political.

same characterization. Instead, the definition says that such mechanisms are simply *apt* to cause similarity in associative thinking. That is, the mechanism might just produce similarity in some of the elements of a characterization. For example, in virtue of being subject to the same psychological convergence mechanism, two individuals might come to believe that jocks tend to be buff, but only one of them, given their unique experience, believes that jocks tend to be bullies.

Similarity of characterization will depend on features of the environment in which individuals are embedded. In some cases, similarly of characterization of two people subject to the same psychological convergence mechanism might be close to wholly overlapping – think of a dictatorship where the citizens are subject to closely monitored and policed 'education,' or propaganda, programmes. In other cases, the similarly in associative thinking will be much less than this. This might be due to all sorts of reasons. One reason could be that aspects of one's characterization might 'withstand' psychological convergence mechanisms. We might acquire some information through education, but nevertheless some of our unique understanding, the elements of a characterization we have developed independent of psychological convergence mechanisms, remains intact. Another reason is that, often, if not always, we are subject to a range of mechanisms that (attempt to) reinforce different characterizations of the same subject matter. This means that individuals internalize different bits and pieces of information from different sources, and thus have different but partially overlapping characterizations.

The degree of similarity that I require is one that is enough to explain coordination of thought and behaviour. For example, our stereotypes about dogs being friendly, or 'man's best friend', is enough to coordinate our thinking about dogs being suited to live in a family home, which undergirds social practices of adopting dogs and caring for them. An exact degree of similarity I will leave open for further research.

One might think: Is it possible that *all* characterizations, or elements of a characterization, are developed through social processes that exert pressure on individuals? Given what has been said above, I don't think so. Sometimes similarity of characterization is not a product social pressure. But further, we all have unique experiences that separate our mental lives from others. Interaction with the world is personal insofar as it is our *own* experience of it. And, this shapes our minds in distinctive ways. Moreover, it is important to recognise that idiosyncrasies in mind-shaping, those elements of one's characterizations that deviate from the characterizations of others, provide unique forms of thought that enables us to challenge dominant ways of thinking, and to reimagine how the social world could be, which subsequently contributes to undermining existing terms of social engagement. I will have more to say about this in chapter two. Finally, it seems that if one were to say that *only* social processes were responsible for the development of characterizations, then this would lead to a regress. After all, if the causal chain of the passing on characterizations were only a function of social processes, then this chain would go back *ad infinitum* – each characterization, or element of a characterization, would be learned from someone else. Thus, there must be characterizations, or elements of a characterization, that are developed outside of social processes, which are then used *in* social processes and passed along to other members of a community. With this in mind, it seems to me that all characterizations will have elements that are more or less idiosyncratic.

Despite this, much of the elements that make up our characterizations are shaped by social processes that exert pressure on individuals, such as education, all forms of public media, desire for conformity, conscious imitation, unconscious social pressure, adhering to the thinking of parents, the effect of news outlets, the testimony of putative experts, those we trust, etc.

To be clear, I will be primarily interested in the elements of a characterization that have been developed as a result of an individual being subject to a psychological convergence mechanism, bolstered by coordination of thought via policies of communication with the same natural language expression, rather than the elements of a characterization that are merely shared or idiosyncratic. And the reason for this is because such mechanisms are an explicit aspect of social reality that can be targeted for resistance and change. And if we are able to take control of these mechanisms, we can reshape how individuals conceive of possibilities for action in the social world.

With respect to conceptual engineering, as I have understood it, we might target characterizations for change since this is apt to bring about changes to our classificatory dispositions; to enact effects on actual social behaviour. Continuing our example of Manne's ameliorative analysis of MISOGYNY, we might think that what is required to get people to change their practices of classifying with 'misogyny' is to change community-wide associative thinking about misogyny, or why we think that the category is important or significant. In particular, the aim is to get people to think of misogyny as a feature of social pernicious environments, rather than the hateful attitudes of bad men.

This brings us to the question asked above: How do we get people to change the characterizations that explain or ground their classificatory dispositions? If psychological convergence mechanisms are responsible for (some) coordination of associative thought, or similarity of characterization, then we can change such characterizations by targeting psychological convergence mechanisms. That is, we do not have to change classificatory dispositions nor psychological information structures directly (say, by changing beliefs). Rather, we can take the more plausible route of making use of, undermining, or else creating convergence mechanisms that help regulate or stabilize similarity of characterization within a community of thinkers and speakers. I won't offer any concrete proposals. This is entirely an empirical matter. Nevertheless, focusing on the environmental mechanisms that maintain overlapping characterizations within a community is a promising location at which conceptual change can be achieved. I will have a little more to say about this in chapter three, §3.7.

## 1.5. Conclusion

A lot has been said so far, so I will sum up. I have stated that my understanding of conceptual engineering will be primarily about *coordinated concepts*. Such concepts are representational traditions within a community. That is, systematically linked networks of individual concepts (i.e., knowledge structures), in which individuals interpret each other as talking about the same thing. Moreover, these coordinated concepts tend to foster the spread of similar *characterizations* and *classificatory dispositions* within this community.

To reiterate, characterizations are bodies of information structured along the dimension prominence and centrality. And, classificatory dispositions involve treating or seeing aspects of the world, on the basis of certain (implicit) beliefs or opinions, as belonging to the same kind. When it comes to social kind concepts, characterizations are in complex interaction with classificatory dispositions. Importantly, characterizations shape our understanding of the importance of significance of a category or kind; and this underlies our tendency to classify one way rather than another. Put differently, our dispositions to construct a characterization is apt to influence the way we classify aspects of the world.

On top of this, I argued that what helps to regulate similarity of characterization are *psychological convergence mechanisms.* Mechanisms of this kind are social; aspects of the socio-epistemic structure of a community that shapes the processes by which cognitive resources are produced and

disseminated; such as the coordinated concepts and overlapping characterizations to which individuals have intersubjective access. Importantly, being subject to the same psychological convergence mechanisms means that members of a community come to have overlapping, or partially overlapping, characterizations.

With this mind, I take conceptual engineering to concern the revision of coordinated concepts insofar as what we are trying to change are the associated classificatory dispositions that make a difference to actual social relations. And one means by which such change is possible is targeting the shared characterizations that are apt to bring about such dispositions – and this might involve undermining or exploiting psychological convergence mechanisms. Put differently, by changing the bodies of information that help shape, guide, or underlie our higher order cognitive competences, we can change our dispositions to classify aspects of the world as belonging to the same kind. But we might wonder: Why would we want to change our classificatory dispositions? Or, to put it differently: What is conceptual engineering supposed to achieve?

# Chapter Two

## Conceptual Engineering: What is it Supposed to Achieve?

*If we are concerned to create just structures, we should be concerned with social meanings.*

Sally Haslanger 2019, p. 19

### 2. Introduction

The previous chapter explored the question: What is conceptual engineering about? I stated that, for my purposes, conceptual engineering is about coordinated concepts: systematically linked knowledge structures associated with, and which underlie, particular classificatory dispositions that emerge within a community of thinkers and speakers. To reiterate, classificatory dispositions involve treating or seeing aspects of the world as belonging to the same (social) kind or category. And, the knowledge structures that interest me are specifically characterizations; bodies of information structured along the dimension of prominence and centrality. Moreover, the characterizations that explain or underlie our classificatory dispositions are often a function of psychological convergence mechanisms that regulate or stabilize coordination of associative thought. This enables the propagation of similar characterizations and classificatory dispositions within a community. On my way thinking, then, conceptual engineering concerns the revision of classificatory dispositions that make a difference to actual social relations. And change can be brought about by altering the characterizations that are apt to bring about such dispositions.

I finished the chapter by asking: What is conceptual engineering supposed to achieve? After all, one might be sceptical that there is any reason to care about our classificatory dispositions being one way rather than another. In this chapter, I will explore reasons why conceptual engineering is taken to be an important form of philosophical theorising, broadly speaking. Specific to my purposes, I will show how revising our classificatory dispositions can play a substantive role in promoting or advancing social justice. But to motivate this idea, I first need to explain the relationship between concepts, ideology, and the construction of social reality. That is, in order to appreciate the social, moral, and political usefulness of conceptual engineering, we must understand that the social world is constructed, in part, through processes involving ideologically oppressive concepts – and that such concepts are locations at which social resistance and change is possible.

### 2.1. Ideology

What is ideology? This is contested. Different theorists use the notion in different ways relative to their purposes. Roughly, what seems to be common to most accounts is that ideology is a representation of social reality that governs the practices of agents embedded in social structure. More precisely, Sally Haslanger argues that ideology is 'the background cognitive and affective frame

that gives actions and reactions meaning within a social system...' (Haslanger 2012, 447).[52] What is in this 'cognitive and affective frame'? According to Eric Swanson, it includes 'a cluster of mutually supporting beliefs, interests, norms, values, practices, institutions, scripts, habits, affective dispositions, and ways of interpreting and interacting with the world' (Swanson forthcoming, p. 6). We can see that there is quite a lot packed into the notion of ideology. And this list might not be exhaustive (Kukla 2018). Nevertheless, it will be sufficient for my purposes. Ideology consists of mutual supporting and overlapping characterizations within a community, that includes beliefs, interests, norms, values, scripts, and the like, which provides the grounds for certain mental and linguistic dispositions; in particular, dispositions to classify aspects of the world as belonging to the same kind or category. And what Swanson draws attention to that I have yet to discuss is the idea that our classificatory dispositions, while grounded in characterizations, are also constitutive of, and supported by, social institutions. For example, we do not just classify certain people as being married on the basis of certain beliefs or opinions. Rather, certain formal institutions, such as government recognition, determine the 'classificatory status' of married people. This institutional anchor provides greater stability to our practices of classification insofar as there can be widespread, and even radical, differences in characterization with respect to the same subject matter across members of a community but nevertheless robustly uniform classificatory dispositions. After all, whether one believes that same-sex marriage is good or bad makes no difference as to whether a same-sex couple is married.[53]

What is necessary to the survival of ideology is that we *enact* it. It is a part of our social practices. What does this mean? How we engage with others and the world around us reinforces and reproduces the cognitive and affective frame that constitutes an ideology. That is, not only does ideological representation 'help to produce and enforce... a *social ontology* that includes specific social identities and relations' (Kukla 2018, p. 2), but rather ideologically infused modes of representation, such as the way we identify ourselves and others, are enactments of ideology and 'keeps it going' (Haslanger 2012, p. 411). Hence, ideology is constantly being reproduced (Kukla 2018, p. 2). For example, when we identify as a particular gender or racial kind, or when we are categorized by others as such, we do not just navigate the social world as someone who belongs to a gender or racial kind, but rather we contribute to sustaining and entrenching an ideology that consists of such kinds.

The enactment of ideology isn't always explicit. It contains 'sneaky' or 'hidden' forms of representation that, nevertheless, manifest in cognitive, affective, and practical dispositions. For example, most people have certain beliefs about animals being food, but the suffering of non-human animals is typically hidden or willfully ignored – to appreciate this, we need only think of the fact that the animals we eat are often labelled in a way that obscures our understanding of their lived existence, such as the terms 'pork' or 'bacon' for pigs (Adams 1990; Podosky 2018).[54] Moreover, ideology often hides behind nature so that we cannot easily see the contingency of social reality. In other words, ideology naturalizes contingent socially constructed facts. Particular characterizations represent the world as being a given, and we are disposed to interact with the world as such

[52] For detailed discussion of recent philosophical work in ideology, see also Mills (1997), Balkin (1998), Stanley (2015), Kukla (2018), and Swanson (forthcoming)

[53] Nevertheless, the systematically linked networks of knowledge structures via natural language signalling strategies, in addition to psychological convergence mechanisms, are able to guarantee a high degree of similarity in characterization and classificatory dispositions *even without* institutional anchoring.

[54] Carol Adams (1990) calls this the *absent referent* – the idea is that animals must be made 'absent' both in existence and in understanding in order for us to consume their flesh. Podosky (2018) believes that this constitutes a distinctive form of *hermeneutical injustice.*

(Haslanger 2019, p. 8, see also Kukla 2018, p. 3).[55] Race is a clear example. Often race is assumed to be a biological category, rather than socially constructed. As a result, we tend to have naturalized understandings of the character and behaviour of certain groups of people. For example, racist science, in particular, often discusses the idea that Black Americans have lower IQs than White Americans in virtue of primarily genetic factors.[56] However, no biological account that would define race as a natural kind has been successful (Zack 1993, 2002; Appiah 1995, 1996).[57] Ideology leads us to believe that the racial categories to which we belong, or ascribed, are given to us by nature, despite being otherwise.

Finally, ideology, as I will understand it, is oppressive.[58] Ideology consists of practical representations 'which stabilizes a particular form of power and domination; or which reconcile and accommodate the mass of the people to their subordinate place in the social formation' (Hall 1996, p. 24 – 25). Specific to my interests, ideology consists of coordinated concepts that enable us to practically orient ourselves in response to material conditions and in relation to others, guiding our behaviour in a way that reliably results in effective and efficient coordination. However, such concepts also play a role in undergirding or supporting oppressive social relations that constitute unjust social hierarchy. This includes racism, sexism, ableism, speciesism, classism, and the like. Put differently, our response to others and material conditions are often configured only to serve the interests of certain groups of people; 'ideology sets us up to participate fluently in practices and structures of injustice' (Haslanger 2019, p. 6). But, exactly how does ideology do this?

### 2.2. Constructing Social Reality

The knowledge structures, or specifically characterizations, that underlie or explain our higher order cognitive processing enables us to render intelligible aspects of the world. That is, we take information from our surroundings as experience, interpret such experience through our characterizations, such as stereotypes or other forms of characterizations, and store this interpretation in long-term memory so that we can access it, in different ways, across a variety of contexts.[59] Put differently, characterizations enable us to make sense of a context, and this undergirds mental and linguistic practices of classifying, categorizing, reasoning, planning, inference-making, metaphorical perspective-taking, and so on.

One way that we make sense of the world is by interpreting aspects of it as having value. Sally Haslanger (2012) calls this a *resource*. And things, as well as people, can be valuable, or resources, in a number of different ways. Something can have religious value, practical value, moral value, epistemic value, etc. For example, we can use different characterizations to interpret figurines as either toys or collector's items; or we can use different characterizations to interpret dogs as either pets or food. How we perceive something as having value depends on the structure of our overall thinking along the dimensions of prominence and centrality – i.e., what features are salient to us, and

[55] Michel Foucault (1980) calls 'hidden' parts of reality, distorted by ideology, *subjugated knowledges*.

[56] Most prominent discussed in the book *The Bell Curve,* which I refrain from citing in full.

[57] There are nearby biological accounts, such as the population naturalism of Kitcher (2007) and Andreasen (2004). Such accounts face their own problems. For example, population naturalism entails that certain groups would count as race when they are putatively not races (e.g., the Amish, Irish Protestants). A further defense of biological accounts of race is offered by Quayshawn Spencer (2019, Ch. 3).

[58] This contrasts with those who think of the notion of ideology as neutral, and only often oppressive.

[59] For others who have formulated the role of representation in this way, see Haslanger (2012, 2015, 2016, 2018, 2019); Fricker (2007); and Camp (2015, p. 601).

how much we believe such features play a role in explaining other features. So, when it comes dogs, if we emphasise their friendliness in our structured thinking practices, this will play a role in our broader understanding of dogs as being suited to live in the family home, or as something to be cared for rather than eaten.

This interpretive practice, when common enough within a community of thinkers and speakers, is one means by which parts of social reality are constructed. That is, when we make sense of our experience by using overlapping characterizations, we participate in the *discursive construction* of shared social reality, which consists of certain norms, practices, objects, cultures, and the like. 'Something is discursively constructed just in cases it is the way it is, to some substantial extent, because of what is attributed (and/or self-attributed) to it' (Haslanger 1995, p. 99). In other words, our dispositions to see the world in a certain way, and to classify it as such, contributes to creating and sustaining parts of social reality. For example, we do not just *recognise* that certain figurines are toys or collector's items. Rather, certain figurines *are* toys or collector's items in virtue of our recognition of them as such. Other examples are clear. Despite being commonly thought to be biological, race is typically theorized within philosophy as a social construction that is produced through discursive processes of classification and naming. That is, on standard accounts, races are discursively constructed through the assignment of social significance to salient biological features and presumed ancestral links to geographical regions (Outlaw 1990; Gooding-Williams 1998; Mills 1998; Haslanger 2000; etc.).[60]

Interpreting aspects of the world as resources, and creating or sustaining their social existence, sets up the possibility for coordinated practice and communication in response to such resources. For example, with the coordinate concept of food, we are able to interpret, recognise, and classify things as food *qua* resource, which subsequently enables us to practically orient ourselves in response to food, such as engaging in more or less complex eating or culinary practices – such as having certain food for breakfast (e.g., cereal), or restricting our diets to comply with moral permissibility (e.g., veganism), or celebrating holidays that centre around the consumption of particular animals (e.g., the Christmas ham), etc.

What's more is that our classificatory dispositions do not just create social phenomena, nor do they simply capture and reinforce what already exists. Rather, how we classify can make a substantive difference to the realm of extant social facts. That is, through communal classificatory dispositions, in particular with social kind terms and concepts, we initiate and sustain *looping effects* that contribute to the alteration of social kinds or categories (Hacking 1995; Mallon 2016).[61] According to Ian Hacking, looping effects involve the following:

> New sorting and theorizing induces changes in self-conception and in behaviour of the people classified. Those changes demand revisions of the classifications and theories, the causal connections, and the expectations. Kinds are modified, revised classifications are formed, and the classified change again, loop upon loop (1995, p. 370).

To put it differently, looping effects are the processes of mutual causal feedback between our practices of classifying people into social kinds and the subsequent alteration or re-construction of

[60] A similar form of construction is said to be the case for gender kinds. See (Haslanger 1995, 2000; Jenkins 2016; Dembroff forthcoming; Barnes forthcoming, etc.).

[61] Hacking (1995) is primarily interested in scientific kinds, but nevertheless his argument can be extended to the construction of social kinds.

such kinds.[62] As Robin Dembroff puts it, 'by developing classificatory practices, we create social kinds, which in turn impact classificatory practices, and so on' (forthcoming, p. 7). Social reality comes to fit what we expect from it because, through our classificatory dispositions, we have made it that way (Wittig 1993, p. 11 – 12).

A quick note before I continue. I take Hacking's (1995) original formulation of looping effects to be too limited. For Hacking, a looping effect requires that the subject of classification take up the classification ascribed or applied to them, and through this uptake the classification changes, and so too does the corresponding kind. I am more permissive than this. Looping effects are, broadly speaking, causal feedback loops between classification and the social world – and this needn't require that one take a term to apply to oneself.[63] As Sally Haslanger argues, 'the causal mechanism of looping need not always proceed through the... identity and intentionality of agents involved' (2017, p. 283).

Given this, looping effects are not just limited to the alteration of social identity categories, as Hacking seems to think. Rather, material reality can be altered through revised classificatory dispositions, that is, amended practices of classifying aspects of the world as belonging to a particular kind – such as individuals pushing back on coordinated classificatory practices and associated characterizations. For example, the growing number of vegans, and those simply interested in more ethical or healthier lifestyles, has seen new practices of classifying with the term 'meat' and 'dairy.' Such terms were previously reserved for products that were derived from non-human animals. Now, even with push back, the terms are regularly applied to plant-based items that bear a resemblance, and play a similar culinary role, in most cases, to real flesh and natural dairy. In other words, our practices of classifying with 'meat' and 'dairy' to plant-based items have played a substantive role in changing the social kinds *meat* and *dairy.* There was a shift in shared characterizations of *meat* and *dairy*, which brought about changes to community-wide classificatory dispositions with the terms 'meat' and 'dairy', and which resulted in the reconstruction of the social categories of *meat* and *dairy.* And, as we can see, no self-understanding on the part of the subject of classification (i.e., meat and dairy) was required for this looping effect to take place.

It's important to be clear about the fact that looping effects lead to alterations in original concepts and categories. For example, our concept FOOD delimits the category of *food*, and our reactions to the items classified as 'food' affects our culinary traditions, tastes, and emotions, associated with FOOD. This revised understanding leads to new characterizations which, in turn, affect shared classificatory dispositions, leading to a revised category *food**. To see this, let's explore a real world case. Our original concept FOOD delimits the category of *food* so as to include products derived from non-human animals. However, within certain communities, there has been a particularly strong reaction to this inclusion – namely, the assessment that it is morally impermissible to eat non-human animals. This has had the effect of changing community-wide associations with FOOD. In particular, it translated into new characterizations of *food*, and subsequently revised classification and eating practices within the community. Specifically, people developed new cognitive and affective dispositions such as believing that animals are not food, or believing that animals are persons, or feeling disgust at the thought of eating animal flesh, etc. And, as a result, people stopped eating

[62] Here I am understanding social kinds as sets that are (roughly) demarcated by coordinated classificatory dispositions.

[63] In this sense, looping effects are a process of *discursive construction.* The difference is that looping effects take place once an initial discursive construction has been established. So, the discursively constructed category *women* can change through revised discursive attributions.

products derived from non-human animals. This affected the category of *food** within certain communities: *food** no longer includes products derived from non-human animals.

In addition to this, alterations to coordinated characterizations and classificatory dispositions can lead to stable changes in how we categorize aspects of the social world, the characterizations that undergird such categorizing dispositions, and associated external practices. For example, the new category of *food** is partly responsible for a shift in material conditions, namely, the proliferation of plant-based foods (i.e., mock meats that are made from wheat, pea, or soy protein). In virtue of this, and the popularity of vegan diets, altered characterizations and classificatory dispositions emerged in order to distinguish between kinds of people who don't eat animal products. There are 'ethical vegans,' whose dietary choices are based on moral considerations; there are 'health vegans,' whose primary concern is with eating healthily; there are 'fashion vegans,' whose decision not to eat animal products is a matter of 'jumping on the band-wagon,' as it were; etc.

Recognition of the looping effects that alter or reconstruct parts of social reality locates a site for the disruption of existing mental, linguistic, and behavioural practices. This is especially important for social justice projects. Change to existing social relations can be brought about by either revising the way we classify aspects of the world or by tinkering with parts of material reality. The former, broadly speaking, is how I have understood conceptual engineering.[64] And though the latter is not conceptual engineering, changing material reality is certainly one means by which change to classificatory dispositions can be achieved. I will have more to say about this in the next chapter.

Apart from discursive construction and looping effects, there are also coordinated concepts that influence how resources are distributed, shaping how we coordinate with others. So, not only do we bring food into existence through dispositions that initiate and sustain looping effects, we distribute such food, through our social structures, on the basis of interpretations that, say, determine the kind of food that is 'fitting' for certain people, which is typically reinforced by material conditions. For instance, junk food, or take-away, is considered to be fitting for working-class families. This is non-accidentally related to the unequal distribution of material wealth that limits food-options for certain people (i.e., low income families can't afford to buy 'healthier' options, or they work longer hours and therefore do not have time or energy to cook, etc.). Because of this association, junk food places tend to be built in lower-income areas, which reinforce preferences through habituation. Thus, not only do our associations play a role in bringing into existence certain facts about social reality, such facts about social reality reinforce and support our associations. Mind and materiality are interdependent and mutually supporting.

Apart from cognitive processing, it's important to make explicit the role of characterizations in regulating *affective* dispositions. This is a core practice of shaping minds (Jones, Schroeter, and Schroeter 2020, p. 86). Affective mind-shaping is closely connected to action or behaviour insofar as it conditions 'cognition and motivation through controlling salience and interpretation and hence rearranging action-options in a hierarchy of possibilities' (ibid. 2019, p. 87). In other words, our decisions to act one way rather than another are constrained by the role of emotional valence, a constitutive part of characterizations, in limiting the options we see as plausible or fitting for the context (e.g., people from low income families will feel that certain food 'for them,' and other food is

[64] Though, this is in a very broad sense. Conceptual engineers often only care about concepts that determine reference or truth-conditions. I am interested in a range of classificatory dispositions and associated patterns of understanding that play a role in constructing social hierarchy. Should this come under the banner conceptual engineering? I think it should. But for anyone who has a strong opinion otherwise, I can simply call it 'psychological engineering.'

not, such as caviar). And this affective regulation is learned. We are taught what to feel. For example, 'a young child behaving badly is told that they are tired rather than angry. A girl, responding negatively to her brother getting a robotics kit for which she has longed while she gets a more gender normative gift, is told that she is jealous rather than indignant at the injustice of it' (ibid. 2019, p. 88). This same jealously is ascribed to, and internalized by, a Person of Colour who wonders why they cannot break into a profession while watching their White counterparts advance past them. This jealousy-attributing move masks injustice, placing the wrong on the Person of Colour (i.e., being jealous) instead of the racially discriminatory systems that keep him out. Thus, affective regulation serves to organize inchoate feelings in one's associative thinking, and this shapes the direction and character of our social practices. Often, the direction to which we are pushed comes at the cost of our wellbeing.

To sum up what's been said so far, how we interpret our experience, through cognitive and affective means, shapes our social interactions. Individuals are disposed to coordinate in response to one another and the material world in accordance with the characterizations and classificatory dispositions that constrain our understanding of possibilities for action, which motivates or guides behaviour. For example, we classify and treat dogs as pets in virtue of particular interpretations that open up possible, or permissible, ways of interacting with them, while closing off others – e.g., we care for dogs, rather than eat them. When this behaviour is copied and reinforced within a community, it becomes a social disposition. We needn't think about having to intentionally treat dogs as pets, instead of food, because we are disposed to see them as pets, and thus we are disposed to treat them as such. Following Sally Haslanger, we can call dispositional patterns of behavior a *social practice*:

> Social practices are patterns of learned behavior that enables us... to coordinate as members of a group in creating, distributing, managing, and eliminating a resource (or multiple resources), due to mutual responsiveness to each other's behavior and the resource(s) in question, as interpreted through [coordinated concepts] (Haslanger 2018, p. 245).

And, again following Haslanger, when social practices are repeated and become entrenched, we can call this a *social relation.* Systems of interdependent social relations we can call *social structure* (Haslanger 2012, 2013, 2015, 2016, 2018, 2019). Thus, in a social structure that sees dogs as pets (i.e., a resource), there are entrenched social practices of grooming, caring, walking, buying, selling, etc. And repeated social practices, or social relations, give rise to further resources that become institutionalized such as grooming businesses, dedicated dog parks, doggy day care centres, pet stores, and the like.

### 2.3. Ideology and Social Construction

What I have said so far does not yet explain the relationship between *ideology* and social construction. To reiterate, the coordinated concepts of an ideology are oppressive. And we can understand this as the idea that such concepts are used to interpret aspects of the world in a way that motivates and undergirds *oppressive* social practices or relations. How to understand this is a difficult question. In one sense, we might say that coordinated concepts are partly *constitutive* of oppressive practices, and thus, constitutive of oppressive social structure. This is might be true for some concepts, such as the characterizations (i.e., pernicious stereotypes) that ground mental and

linguistic dispositions with respect to *women*, and which undergirds oppressive practices, such as constraining women's agency in a way that (unjustly) denies life opportunities (e.g., access to high-paying jobs). However, this might not be true for all coordinated concepts. Some might simply be causally efficacious in bringing about oppressive practices. For instance, owing to repeated associations between our concept of urban youth and negatively valance associative thinking about Black Americans, the term invokes a socially reinforced characterization that feeds into oppressive practices. With this said, however, there might not be such a great difference between the coordinated concepts that are partly constitutive of oppressive practices, and those that are merely causally efficacious. I do not wish to get into a debate over this difference. All that needs to be said is that coordinated concepts *contribute*, whether causally or constitutively, to oppressive social practices.

What is oppression? There are a few things to note. Broadly speaking, oppression can come in two forms. First, oppression can be *repressive* insofar as individuals are directly coerced by an overarching authority (Haslanger 2019, p. 5).[65] Second, oppression can be *ideological* insofar as it is 'enacted unthinkingly or even unwillingly by the subordinated and/or privileged' (ibid. 2019, p. 5). After all, ideology provides frames of thought, which are responsible for governing mental, linguistic, behavioral dispositions in a way that subordinates some and privileges others. My interest is in ideological oppression. Of course, repressive oppression is a serious form of injustice. However, it is widely available for all to see and judge. We can look at a fascist regime and understand very quickly that things should be different. Ideological oppression, on the other hand, is a form of oppression that tends to operate below the level of conscious awareness; it is typically hidden from individual and public understanding:[66]

> It is embedded in forms of coordination that we are highly motivated to engage in, and engage in fluently; results of such coordination are often valuable; and the terms of such engagement come to constitute our self-understandings and our identities (ibid. 2019, p. 6).

For example, our interpretation of animals as food, which is deeply embedded in our culinary institutions, undergirds a network of social relations built on the consumption of animals that we, as a community, willingly engage in; and this does not just include a desire to eat animals on the basis of their aesthetic appeal or nourishment, but also to participate in cultural practices of going out to dinner with friends and family, or exercising a form of masculinity through the consumption of red meat, or celebrating holidays that are centered on eating animals, etc.

Linking these thoughts to the foregoing discussion, ideological oppression concerns the unjust social practices and relations that people enact that rely on the internalization of particular coordinated concepts. And, as was discussed in chapter one, this process of internalization is often a matter of members of a community being subject to psychological convergence mechanisms that regulate or stabilize coordination of associative thought, and this underlies or explains communal classificatory dispositions. That is, ideological oppression gets a hold of us through, *inter alia,* the internalization of characterizations, which undergirds practices of classifying, made possible through

[65] I leave open the whether 'legitimate' authorities can be repressive insofar as, say, a democratically elected leader of a country can exercise unjust power.

[66] Of course, repressive and ideological oppression often come hand-in-hand. People might see repressive oppressive as justified in virtue of being in the grips of an ideology.

environmental mechanisms of psychological convergence within a community (e.g., education system, adhering to the thinking of parents, etc.).

To be explicit, let me lay out the relationship, as I see it, between the mechanisms that regulate associative thinking within a community and the creation or maintenance of unjust social structure. (1) We internalize pernicious coordinated concepts via psychological convergence mechanisms that regulate associative thinking within a community;[67] (2) this regulated associative thinking within a community (at least partly) underlies or explains our shared classificatory dispositions; (3) our shared classificatory dispositions create and sustain a social ontology, which is reinforced by looping effects; (4) this social ontology consists of resources (e.g. food) that members of a community coordinate their social behaviour in response to; (5) and when this coordinated social behaviour is repeated and entrenched, it constitutes an aspect of social structure. (6) Finally, the presence of social structure reinforces pernicious coordinated concepts, which contributes to the process of internalization of such concepts, through ideological mechanisms such as the distribution of material resources, and the appearance of being natural. To see how this process translates into structural injustice, consider an example offered by Sally Haslanger (2016), that she, following others, calls *The Invisible Foot*:

> Imagine a couple, Larry and Lisa, who, we suppose, are equally intelligent, talented, educated, and experienced in the workplace; they have equal power in their relationship, have no prejudices about gender roles, and are equally capable of all domestic tasks and childrearing tasks. Larry and Lisa decide to have children; baby Lulu arrives. They live in a community where decent childcare is beyond their means. Moreover, let's suppose that in this community, as elsewhere, there is a wage gap: women, on average, make only 75% of what men make. Under these conditions, unless Larry and Lisa have special reasons to think that they are unusual in their earning capacities, it is reasonable for Larry to work full-time and for Lisa to make adjustments in her work, e.g., to work part-time, to take time off, to take a less demanding job. But in our society, "wealth determines power, domestic work is unpaid, and divorce laws do not evenly divide wealth" (Cudd 2006, p. 149). So Larry accrues greater human capital and ends up with more power in the relationship. Moreover, insofar as Larry and Lisa are typical, women on average will be poorer risks for employers who will "tend not to trust that women will stay with their careers or that if they do, they will devote the kind of time and energy to them that men will" (Cudd 2006, p. 149). As a result, "women's jobs" that require less commitment, mobility, and experience will pay less, and women will have to prove themselves exceptional to be considered for high paying "men's jobs." As a result, a pattern is reinforced (Haslanger 2016, p. 122 – 123).

This example shows a form of structural injustice against women. It does not appeal to features of the individuals or their capacities, but rather their embeddedness in a social, gendered environment. In particular, it reveals how action-options are constrained to the extent that what is reasonable or rational for one to do in a situation can contribute to one's own subordination or oppression.

How does this relate to the foregoing discussion? The actions available to Larry and Lisa differ in virtue of the interpretive practices and classificatory dispositions of the community of thinkers and speakers in which they are embedded. Plausibly, the reason for this difference is because in our society men are understood to be 'breadwinners,' and as such they are expected to participate in social structure in a way that confers them power and status. Larry can work, accrue wealth, gain

[67] As well as participating in natural language signaling strategies that help foster the spread of similar characterizations and classificatory dispositions within a community.

status, and therefore amass power. This is not available to women, like Lisa, who are understood to be 'caretakers,' and treated as such. This interpretive practice, a result of community-wide characterizations that underlies or explains communal classificatory dispositions, means that working high-paying jobs is less of a possibility, and therefore so is wealth, status, and power. And part of the reason for this is that such differences in opportunities for work owes to ideological mechanisms such as prevailing beliefs or opinions within a community about the *nature* of men and women – it is natural that men go to work and women take care of the children, or so the thought goes. What I hope to have highlighted is that coordinated concepts with their associated characterizations play a substantive role in creating and sustaining a social structure in which some (e.g., men) are privileged, and others (e.g., women) are subordinated.[68]

The process that I have outlined is not the only story of social construction. Rather, it is illustrative of the kinds of steps that are often involved. Part of the reason for saying this is because we enter into a world that already exists. So, much of the classificatory dispositions that we perform do not create resources, such as 'breadwinners' and 'caretakers,' but instead sustain them. Importantly, the internalization of coordinated concepts is often not *prior* to social ontology – that is, social ontology, the things that exist in a particular social world, supports the process of internalization owing to the fact that merely seeing the world as such plays a role in one classifying the world as such. In other words, we see the world as consisting of objects with certain properties, or observed regularities, and this shapes our expectations that play a normative role in reinforcing those observed regularities. Moreover, other cognitive and affective processes will be involved in the process of construction – affect, in particular, is a core people-making practice (Jones, Schroeter, and Schroeter 2020) Further, there will be other kinds of ideological mechanisms that play a role in reinforcing pernicious ways of seeing. I have only listed two that strike me as particularly important.

## 2.4. What is Conceptual Engineering Supposed to Achieve?

With the relationship between concepts, ideology, and social construction in place, let's return to the question stated at the beginning of this chapter: What is conceptual engineering supposed to achieve? Or, perhaps to put the question in more loaded terms: Why care about conceptual engineering?

One might think that conceptual engineering is just about achieving or bringing about *meaning* change. Perhaps the most prominent theorist who defends this way of thinking is Herman Cappelen (2018). He argues that conceptual engineering 'should be seen as having its *goal* to change the extensions and intensions of expressions' (ibid. 2018, p. 61 my emphasis). So, when someone like

[68] Of course, intersectionality complicates this rather simplistic story (e.g., Crenshaw 1991; Patricia Hill-Collins 2003; Yuval-Davis 2006). Briefly, intersectionality is the idea that the many different social categories to which we belong intersect and thus produce unique social locations that involve distinctive forms of privilege and subordination. This intersection is not merely the addition of one's social categories. Rather, it is a complex and often unpredictable relationship that generates unique social experience. Given this, there will be cases where some men will be worse off in society than some women. For example, Indigenous men in Australia are subject to severe forms of oppression owing to on-going colonization to the extent that they are plausibly worse off than White women. A similar story might be said with respect to trans* men and cis women. Other examples abound. Unfortunately, I do not have the space to go into the complex details of how the social categories to which we belong sit in an irreducible and distinctive relationship. Nevertheless, I want to make explicit that I am committed to intersectionality; in particular, a type of intersectionality that is non-additive.

Sally Haslanger offers an ameliorative definition of 'woman,' we can understand the primary goal of her project as changing the extension or intention of 'woman.'[69]

Despite this, changing meaning isn't the *only* goal available to conceptual engineers. In fact, it really isn't the main goal of, at least many, ameliorative projects. For most, the goal is to enact practical effects – to bring about real world changes to our thinking, speaking, and behavioural practices. An oft-quoted passage from Burgess and Plunkett sums this up clearly:

> Arguably, our conceptual repertoire determines not only what beliefs we can have but also what hypotheses we can entertain, what desires we can form, what plans we can make on the basis of such mental states, and accordingly constrain what we can hope to accomplish in the world. Representation enables action, from the most sophisticated scientific research, to the most mundane household task. It influences our options within social/political institutions and even helps determine which institutions are so much as thinkable. Our social roles, in turn, determine what kind of people we can be, what sorts of lives we can lead (2013, p. 1096).[70]

To put it differently: our engagement with the world depends on how we understand it. One means by which we do this is through our representational devices that sets limits on possibilities for intentional action, and which constrains and enables us by establishing social kinds for us to adopt and recognise in others (e.g., gender and race concepts), as well as shaping how we understand and interact with material reality (e.g., what things count as falling under 'food'). In other words, we organise our behaviour on the basis of coordinated concepts that are used to narrow the range of potential ways of practically orienting ourselves in relation to others and the world around us.

With this in mind, a better way of understanding Haslanger's project, one more in line with her way of thinking, is this: Haslanger is not *just* advocating for a change in extension and intension to our gender terms or concepts. Rather, the change in meaning is meant to serve certain social, moral, and political purposes. Namely, to advance pursuits in the project of social justice. The goal is not just to change meaning, but for such change to bring about better social relations.

It's important to note, however, that meaning change is not the only means through which desired or intended practical effects can be achieved. To reiterate what was said in the previous chapter, it is possible that meaning (i.e. extensions and intensions) can come apart from our mental and linguistic dispositions. Put differently, *The Standard View* is not the only way to think about conceptual engineering. Moreover, meaning change *might not even bring about practical effects at all*. We can 'change the meanings of words used by people without many of them noticing and changing their [mental and] linguistic practice accordingly' (Koch forthcoming, p. 4).[71] So, to be explicit: my understanding of what conceptual engineering is supposed to achieve are certain practical effects, such as changes to mental, linguistic, and behavioural dispositions, which feed into broader social relations, and this needn't be brought about as a result of changes to the meaning of our terms or concepts. Meaning change, or a change in extension and intension, if we're lucky, is epiphenomenal:

---

[69] Accounts of conceptual engineering that argue that the goal of conceptual engineering is meaning change are abundant. Here is a list of examples: Haslanger (2000, 2020), Thomasson (2020), Ball (2020), Richard (2020), Pinder (forthcoming), etc.

[70] Though, it should be noted that Burgess and Plunkett (2013) are not talking about characterizations in this passage.

[71] And even if we *can't* change meaning without also changing classificatory dispositions, it doesn't matter. We've got reason to be interested in these dispositions however they do (or don't) relate to meaning.

it is an extra layer that comes about as a result of the kind of change that interests me – change to actual mental, linguistic, and behavioural dispositions.

Which range of practical effects does conceptual engineering intend to bring about? Like Haslanger (2000, 2020), as well as Dembroff (2016, 2020), Jenkins (2016), Manne (2018), Barnes (2016, 2020), and others, my concern is with the role of conceptual engineering in promoting and advancing social justice. So, to be more specific about what conceptual engineering of social kind concepts is supposed to achieve, my answer is this: the goal of conceptual engineering is to contribute to bringing about more just social relations; to advance pursuits in the project of social justice. And the means by which it can accomplish this is through, *inter alia*, revising our classificatory dispositions; the way we treat or see aspects of the world, on the basis of (implicit) beliefs, opinions, or other attitudes, as belonging to the same kind.[72] In other words, we must undermine the role that coordinated concepts play in sustaining shared ideology that shapes current norms of social engagement. Unjust social structure can be undone, or at least resisted, through changes to our mental and linguistic dispositions.[73] After all, our shared characterizations and classificatory dispositions play a substantive role in the creation of social structure. Without coordinated concepts, we would not be able to create and sustain the relations that constitute social structure. Thus, in order to combat structural injustice, we can target the coordinated concepts that undergird oppressive behavioural practices. We can change our characterizations, and the classificatory dispositions that such characterizations underlie or explain. In turn, this will translate into better social practices.

Moreover, achieving social justice is also a matter of ensuring that we have the right interpretive practices and social ontology. For example, *hermeneutical injustice*, a species of epistemic injustice, occurs when a marginalized person is unable to render intelligible experience, and therefore unable to convey this experience across social space, owing to a lacuna in a collective set of cognitive resources, which includes classificatory dispositions and characterizations (Fricker 2007, Ch. 7). This occurs under conditions of hermeneutical marginalization: prejudicial constraints that prevent marginalized groups from participating equally in the structural processes that give rise to the shared cognitive resources of a thinking and speaking community. So, for example, prior to the concept of bisexuality circulating within our community, people who were sexually attracted to both men and

[72] Jennifer Nado (2020, m.s.) has argued for a similar position about *some* projects in conceptual engineering. That is, conceptual engineering aims to change classificatory procedures in order to achieve certain practical ends or to fill a practical role. And one role, importantly, is advancing social justice. The main difference between Nado's version of conceptual engineering and the one that I have offered is that Nado believes that conceptual engineering involves engineering *abstract* entities – that is, the engineering of classification *procedures* rather than *dispositions.* In fact, Nado even suggests that the engineering of dispositions does not fall within the purview of conceptual engineering. So, to put it differently, the terminus of Nado's conceptual engineering projects lie in the stipulation of a procedure for classifying things as falling under the same kind or category. This strikes me as odd. Many of the arguments she offers for shifting our understanding of conceptual engineering to classification rather than meaning seem to focus on how revisions to the former will bring about actual change to mental and linguistic behaviour. However, if classificatory procedures are abstract, then there is no guarantee that engineering such things will bring about the kind of change that she seems to be interested in – after all, what good are revised abstracta that have no causal influence over our dispositions? This gives us at least prima facie reason to believe that while the revision of classification procedures is important, the main goal of conceptual engineering is changing classification dispositions, perhaps among other things.

[73] Sometimes what justice might require is first changing a social practice, and then changes to thought follow. As a conceptual engineer, I am mostly interested in the other direction: changing thought which is apt to cause changes to social practice.

women were not able to make sense of themselves as having a particular sexual identity; an understanding that is in their interest to have. Conceptual engineering is a vital means of ensuring that hermeneutical injustice can be adequately overcome. After all, we must work out which are the best coordinated concepts to fill in a lacuna in the collective set of resources that will enable marginalized people to appropriately render intelligible their experience – and, of course, such cognitive resources must be developed by, or in coordination with, those who are situated within non-dominant situatedness. I will say more about hermeneutical injustice, and related injustices, in chapters five and six.

Similar to this, Robin Dembroff (forthcoming) has recently introduced the notion of *ontological oppression.* This occurs when a social kind has oppressive membership conditions, or alternatively, when there are social contexts that unjustly fail to provide membership conditions that would construct a social kind. Specific to gender, Dembroff argues that 'genders that ought to be recognized may not be, and there may be recognized genders with unjust membership conditions' (forthcoming, p. 2). In a sense, ontological oppression is the metaphysical correlate of hermeneutical injustice. The difference, of course, is that where the wrong of hermeneutical injustice is located at the level of understanding, the wrong of ontological oppression is located at the level of existence – what kinds there are, or whether there are kinds at all. So, to go back to the example above, ontological oppression was present when there were no social kinds available for people who were sexually attracted to both men and women to fall under, or identify with. Like hermeneutical injustice, conceptual engineering will be an important means of overcoming ontological oppression. This is because through creating new classificatory practices, or else revising existing ones, we can bring about new social facts – either by creating novel kinds, or by altering the kinds we have through looping effects. That is, conceptual engineering, understood as the project of developing new or revised forms of classification, allows us to examine which kinds we need, and how such kinds should be constructed.

There are many different conceptual engineering projects, with social justice in mind, that can be understood as attempting to either fill gaps in the existing set of cognitive resources or provide the social world with either new or revised social kinds or categories. Most prominently are the projects that aim to make changes to our dispositions with 'woman' (and 'man') so as to be (more) trans* inclusive. One of the aims of such projects might be meaning change, but nevertheless the important social justice work will involve how it is that trans* people are in fact treated. Importantly, how we classify trans* people is a question of justice – it seems reasonably clear that we should not misgender or dead-name trans* people (Dembroff and Wodak 2018).[74]

Further, appropriate classification of what exists is not (always) the end goal. To reiterate, how we classify aspects of the world makes a substantive difference to extant social facts. Through looping effects we can alter or reconstruct social phenomena. For those social justice projects that aim to achieve trans* inclusivity, one goal might be to ensure that we explicate a concept of woman that captures the fact that trans* women are indeed women (e.g., Jenkins 2016). However, another project of social justice surrounding trans* inclusivity is to accept that even *if* trans* women are not women, they *should* be (e.g., Dembroff forthcoming; Barnes 2020). The goal is to bring about a world in which trans* women are women, and this end requires the process of looping certain people

[74] And, there is at least *prima facie* reason to think that respecting the classification of trans* people, or simply respecting trans* people themselves, will make the lives of trans* people much better. Plausibly, it will contribute to decreasing the comparatively high rates in suicide, attempted suicide, and homicide – especially of trans* people of colour.

into a social category – revising our characterization(s) of women such that this will bring about revised classificatory practices that alters social category *woman* to be trans* inclusive. Moreover, part of this project might even require refusing to classify people on the basis of the kind to which they actually belong (Dembroff forthcoming). Instead we might be required to classify on the basis of the categories that we want, or that will achieve social justice.[75]

## 2.5 Conclusion

Let me sum up the core point of this chapter. Conceptual engineering, for the purposes of this thesis, is supposed to achieve a world in which there are more just social relations. And it can achieve this by (i) *altering* existing characterizations that will underlie revised classificatory dispositions which make better sense of values and interests (e.g., revising dominant biological understanding of women to motivated new practices of classifying with 'woman'); (ii) *appealing* to existing characterizations to argue for revised classificatory dispositions that make better sense of values and interests (e.g., appealing to existing characterizations of woman to argue that trans* women should be classified as 'women'); (iii) *revising* existing classificatory dispositions in order to change characterizations, which will in turn reinforce the new classificatory dispositions via looping effects (e.g., changing our classificatory practices with 'women' to apply to trans* women, which will alter our understanding of the significance of importance of the category *woman*); or (iv) *bringing in* new coordinated classificatory practices *ex nihilo* to meet unmet hermeneutical needs (e.g., introducing classificatory practices with 'non-binary' to make sense of particular gender non-conforming people).

The foregoing options raise an important question: how might we bring about conceptual change? This has come to be known as *the implementation challenge of conceptual engineering.* Roughly, the challenge concerns how the concepts that we design in theory can take root propagate within a particular context (i.e., speech situations, communities of thinkers and speakers, etc.). In the next chapter, I will explore this challenge, and argue that existing formulations are posed too individualistically. In order to see just how hard the implementation challenge is, we must pay attention to the social infrastructure and forces that play a substantive role in stabilizing dominant terms of conceptual engagement, and which stifle conceptual innovation.

[75] Another clear example of bringing certain social categories into existence for social justice purposes is bringing into existence the gender kind *non-binary.*

# Chapter Three

## Implementation and Reproduction

*The revisionary analyst drops a mutation in to a population, hoping that it will "reproduce" and in one way or another establish itself, even replace all of its alternatives over time.*

Mark Richard 2020, p. 377.

### 3. Introduction

In the foregoing chapter, I offered an account of the goal of conceptual engineering with respect to social kind concepts; or what conceptual engineering is supposed to achieve. To reiterate, conceptual engineering, as I have understood it, is supposed to contribute to achieving a world in which there are more just social relations. Importantly, we can bring about such a world, through conceptual engineering, by revising our classificatory dispositions. This is because our practices of classifying play a substantive role in the creation and maintenance of social reality. Through classification, we can appropriately make sense of existing social kinds or categories; we can alter existing social kinds or categories through looping effects; or we can bring into existence new social kinds or categories *ex nihilo*. Each of the foregoing can contribute to aiding in the process of achieving social justice. In other words, what justice often requires is, *inter alia*, bringing about alternative ways of understanding the world (e.g. Fricker 2007), reconstructing aspects of social reality, or else introducing new kinds of people and material resources (e.g. Dembroff forthcoming).

With this in mind, we might wonder: conceptual engineering seems all well and good in theory, but how would we actually go about getting people to adopt a proposed conceptual practice? This has come to be known as *the implementation challenge of conceptual engineering.* To date, there has not been a proper formulation of this problem. Though some have offered rough ways framing the challenge. Of course, different formulations are suggested within different frameworks of conceptual engineering; many of which focus on how meaning change can be achieved, or whether it can be achieved at all (e.g. Cappelen 2018; Deustch 2020). Ultimately, I will argue that the implementation challenge of conceptual engineering is often posed too *individualistically*. What does this mean? Briefly, the implementation challenge is framed in a way that focuses too strongly on the role of *individuals* in persuading or convincing others in their community to take up an alternative conceptual practice.

In this chapter, I will cast doubt on individualistic implementation strategies, or strategies for implementing concepts that focus mostly on individual conceptual advocacy. I will argue that such strategies ignore or fail to appropriately take into account the social infrastructure (partly) responsible for maintaining the conceptual status quo, or which prevent individuals in the grips of an ideology from responding positively to ameliorative proposals that would promote justice. Put differently, I will make explicit the entrenched structural systems that prevent conceptual change; the social forces that keep things conceptually as they are. Specifically, I will discuss a particular social process that regulates community-wide classificatory dispositions. That is, I will introduce and

examine the notion of *conceptual reproduction.*[76] Roughly, this refers to the practice of individuals non-accidentally copying prior stable patterns of classification within a social and representational milieu. And importantly, there are different social mechanisms and incentives that play a substantive role in pushing a conceptual practice forward through reproduction.

It should be made clear that understanding conceptual reproduction is important for conceptual engineering. The latter is about intended changes to a conceptual repertoire; the former concerns how an existing conceptual repertoire is maintained. Thus, conceptual engineering and conceptual reproduction stand in tension with one another. The plans of conceptual engineers to implement an improvement to a conceptual repertoire will have to negotiate the mechanisms that sustain extant classificatory practices. Given this, part of the implementation challenge is discerning how to either work with or against conceptual reproduction. It is a matter of discovering ways to exploit or disrupt reproductive mechanisms in order to improve the likelihood of stably refining or revising classificatory dispositions within a community.

The upshot of bringing to light conceptual reproductive mechanisms is that we can begin to see that individualistic implementation strategies will struggle to compete with the social forces that tend to favour dominant terms of conceptual engagement. This should make us sceptical that individualistic strategies, on their own, are able to successfully bring about changes to the social kind concepts operative within a community. Given this, we need multiple implementation strategies, some of which go beyond individual conceptual advocacy, and that are sensitive to facts about the structural barriers that stifle conceptual innovation. Put more roundly, it is only until we see how much social infrastructure is in place to maintain the conceptual status quo that we will come to appreciate just how powerless individuals can be in their effort to bring about conceptual change.

### 3.1. What is the Implementation Challenge?

In order for conceptual engineering to achieve its goal, whatever it might be, there must be a means by which the concepts we develop in the armchair can be implemented out in the wild. Otherwise, despite appearing to be a promising philosophical methodology in theory, conceptual engineering will be useless for our practical lives. This gets to the heart of a problem known as *the implementation challenge of conceptual engineering.* To reiterate, there isn't yet a proper formulation of this problem. Though some have offered approximations in passing. For example, Herman Cappelen and David Plunkett claim the following:

> Once you have settled on an ameliorative strategy, you might want to do some work to implement it, that is, you might want to engage in a bit of activism on behalf of your ameliorative strategy. If

[76] There is a sense in which conceptual reproduction, which involves social processes of copying classificatory disposition, also works to regulate shared characterizations within a community. This is because characterizations are in complex interaction with our classificatory practices: characterizations underlie or partly explain our dispositions to classify one way rather than another. So, by stabilizing our classificatory dispositions through mechanisms of social copying, we *indirectly* stabilize the characterizations that underlie or explain such dispositions. This, of course, looks as if I am confusing the causal order. So, to be clear, I am not saying that classificatory dispositions underlie or explain our characterizations. All I mean to say is that when our classificatory dispositions are stabilized, there is a sense in which this gives us reason to think, or have confidence, that our characterizations of a subject matter are apt or appropriate. Given this, conceptual reproduction is a kind of indirect psychological convergence mechanism.

> that's something you want to do, it raises an 'implementation challenge': how are ameliorative strategies best implemented? (2020, p. 3). [77]

Alexis Burgess and David Plunkett offer something similar:

> There is what we can call attempts at conceptual implementation. Roughly, this involves a kind of advocacy, in which one tries to get some people (ranging from a lone individual to a large population), to actually take up and use the concepts one is in favor of (2020, p. 5).[78]

Each of these passages tells us that what the implementation challenge requires of a conceptual engineer, or someone fond of an ameliorative proposal, is to undertake certain forms of activism or advocacy on behalf of their preferred concept. In other words, the challenge of implementation is a matter of what one can do to persuade others in their community to adopt an alternative conceptual practice.

In this chapter, I will argue that framing the implementation challenge in this way warps our perception of the shape, size, and nature of the problem.[79] In particular, it is posed far too individualistically. It focuses too squarely on the role of individuals in bringing about conceptual change without saying much about the social forces that maintain the conceptual status quo, making individual conceptual advocacy much less likely to be successful.[80] Before I spell out this concern in more detail, I will spend a little time getting clearer on what exactly is meant by 'implementation.'

Broadly construed, to implement something is to put something into effect. The 'something' in question differs depending on one's preferred theory of what conceptual engineering is about. Many theories in recent literature focus on reference change (e.g., Cappelen 2018; Haslanger 2020). That is, changing the range of things that fall under a particular kind. Achieving this, the story goes, requires changing the intension of a term or concept. Other theories, just as prominent, require changes to the function, or one of the functions, of terms or concepts, where 'function' can be understood in different ways: etiological or proper function (Simion and Kelp 2020); the purpose or

---

[77] Mark Richard calls this an *A-project*: 'a project of offering and trying to get others to accept a revisionary ameliorative analysis of a concept…' (2020, p. 360).

[78] Other theorists have put forward suggestions as to how one should think about the implementation problem, but nevertheless bring in assumptions about the nature of the problem, such as focusing on meanings within a metasemantic externalist framework (e.g., Jorem forthcoming). I use these quotes because they are reasonably neutral with respect to what a proposed alternative concept might look like.

[79] When it comes to meaning change in particular, Herman Cappelen (2018) has argued that on a moderate semantic externalism, meaning is inscrutable, given that intension and extensions do not supervene only on internal mental states – allowing for the possibility that we might be mistaken or in error in thinking that a certain predicate or term applies to some aspect of the world. Because of this, we lack control over the processes required to undertake conceptual engineering out in the wild. Despite this, argues Cappelen, we should still keep trying to implement engineering proposal since that is the kind of creatures we are (Cappelen 2018, p. 73). Cappelen further thinks that inscrutability, and thus lack of control, is not just a problem for the externalist. Rather, the internalist, one who thinks that meaning does supervene on some set of internal mental states, faces the problem of discerning whether such internal states are under our control (Cappelen 2018, p. 82). See Joey Pollock (forthcoming) for an internalist response this objection.

[80] I accept that social forces are less likely to get in the way of changing purely theoretical concepts, such as TRUTH, EXPLICATION, BELIEF, etc. However, there will still be some forces, often power-affected, that keep theoretical concepts, such as those that are operative in philosophy, that keep such concepts in place (e.g. paradigm philosophical programmes that are associated with prestige, prominent senior figures in philosophy that juniors scholars want to emulate, etc.). Despite this, my interest is primarily in social kind concepts, and such concepts are greatly regulated and protected by social forces.

goal of a concept (Nado 2019; Thomasson 2020); the function of a concept within a system (Haslanger 2020); etc. I won't doubt the plausibility of any of these theories.[81] Rather, I will simply state that on my way of thinking, the relevant effect that must be brought about is change to coordinated concepts. This can involve change to coordinated and stable characterizations within a community, or the communal classificatory dispositions that are associated with such characterizations. The goal, however, is to ultimately bring about change to actual mental and linguistic behaviour – specifically, community-wide practices of classification.

When it comes to implementation, there are not just differences with respect to what conceptual engineers intend to put into effect. There are also differences with respect to what conceptual engineers take to be the scope of implementation. For example, some might only be interested in changing concepts for the purposes of theoretical enquiry (e.g. Scharp 2013), some might be interested in widespread change for particular concepts that operate within entire communities (e.g. Haslanger 2000, Manne 2018); and, even more narrowly, some might be interested in conceptual change simply in a context, such as the stipulation and uptake of speaker meanings (e.g. Pinder forthcoming a, forthcoming b), or the modulation of concepts within conversation (e.g. Jorem forthcoming).[82] I am interested in changes to the coordinated concepts that partly constitute a public language; the shared concepts that cut across subsets of social groups within a linguistic community.[83]

What does it mean to put an ameliorative strategy into effect within a community of thinkers and speakers? Roughly, an ameliorative strategy can be put into effect in four ways: *introduction, elimination, revision,* and *replacement.*[84]

*Introduction* involves the creation of a concept, and instituting this creation within an existing repertoire of coordinated concepts.[85] *Elimination,* on the other hand, is conceptual engineering as demolition. It involves removing a coordinated concept from our existing repertoire.[86] Surprisingly little has been said about this form of conceptual engineering – and it is an important form given the social and political aims of certain ameliorative theorists who wish to trash certain concepts, such as racial eliminativist who believe that racial concepts do not correspond to anything in the world (Appiah 1995; Zack 2002).[87] *Revision* is, perhaps, paradigm conceptual engineering.[88] It is the re-engineering of extant concepts, with the aim to improve it. *Replacement* is the final version. It is about

[81] Some, however, do doubt the plausibility of semantic implementation, such as Max Deustch (2020). He argues that 'on *any* metasemantic view that requires more, for an existing term to have a particular semantic meaning and reference, than just the *intention* on the part of some group of speakers to *use* the relevant term *as if* it had that vert semantic meaning and reference,' (ibid. 2020, p 19) then semantic change via stipulation will be virtually impossible. This is because we are completely in the dark about what else we would need to change than just the intention of speakers.

[82] Of course, the goal of such might be for such meanings to be passed along from context to context and then to fully take root in an entire community.

[83] In the literature on epistemic injustice, this is sometimes referred to as a 'collective hermeneutical resource,' though such resources include more than just concepts.

[84] Matthew Lindauer (2020) has argued that conceptual preservation is also a form of conceptual engineering. That is, conceptual engineers often have good reason to preserve existing concepts – such as when our existing concepts are more 'fruitful' for certain forms of enquiry rather than proposed alternatives.

[85] David Chalmers (forthcoming) call this *de novo* conceptual engineering; Mona Simion and Simon Kelp (2020) call this, one part of, *conceptual innovation*.

[86] Manuel Gustavo Isaac (2020) calls this an *eliminative strategy*.

[87] It's worth noting that neither Appiah nor Zack take themselves to be conceptual engineers, or engaging in conceptual engineering. Despite this, it is clear that they are doing some version of ameliorative enquiry.

[88] Simion and Kelp (2020) call this *conceptual change engineering.*

creating a new concept to swap out with another that is relevantly similar, but not the same. It should be noted that each of these projects is not separated by sharp boundaries. Often it will be unclear whether one is revising a concept or simply replacing it. This partially depends on one's preferred theory of concept identity – on my account, as was discussed in chapter one, it is a matter of being a part of a particular representational tradition.[89]

Of course, which strategies will work for implementation depends on the kinds of concept one is trying to implement. To be clear, I am concerned with those concepts that regulate aspects of our social life: social kind concepts. With this in mind, the task of this thesis is to cast doubt on the plausibility of conceptual implementation strategies, for social kind concepts, that focus mostly on the role of individual conceptual advocacy. Specifically, in the following chapters (four through eight), I will argue against a particular individualistic implementation strategy that has gained some popularity in recent conceptual engineering literature: engaging in metalinguistic disagreement. My aim is to show that this particular strategy is unlikely to succeed, at least on its own, insofar as individuals will struggle to compete with the social forces that maintain dominant terms of conceptual engagement. Such forces erect structural barriers to individuals in their efforts to convince or persuade others within their community to take up alternative classificatory practices (or downgrade their confidence in existing practices). In order to get to this conclusion, I must first explore, in this chapter, some of the social infrastructure that is (partly) responsible for entrenching the conceptual status quo. In particular, I will explore mechanisms of conceptual reproduction.

## 3.2. What is Conceptual Reproduction?

The notion of reproduction roughly amounts to the process by which an instance of some kind (or property, or individual) is produced or copied from an earlier instance of that very kind in a particular environment (e.g., an organism produces from parents). This process explains the continuation of a kind – after all, were the reproduction of a kind to stop, it would become extinct. At its core, *conceptual* reproduction is similar. It is a process by which a conceptual practice, certain dispositions with a term, is sustained and maintained through the production or copying of earlier instances of that practice within a community.

To be clear, what are reproduced are tokens of a *word* associated with particular cognitive dispositions. Specifically, dispositions to classify aspects of the world as belonging to the same kind. And the reproduction of a word-token is a forward causal relationship of copying earlier conceptual practices. In other words, conceptual reproduction occurs within a community in which members copy prior stable patterns of classification.[90] For example, we classify aspects of the world with 'food' because, at least in part, we sit in a causal relationship of token copying with past speakers. And the force of precedent moves the practice forward. By copying earlier tokens, our classificatory practices unfold into the future. If such tokens were not copied, then the conceptual practice would become extinct.[91]

[89] It's worth noting that there may be vague cases of concept identity.

[90] Of course, there is an element of induction involved. We don't have direct access to others' dispositions, but we observe their classifications and our attention is shaped by background conversational context – and this shapes our understanding of the relevant similarity class picked out.

[91] This is not to say that the concept wouldn't exist at all. Perhaps it does in some abstract sense, such as Nado's (2020) classificatory procedures. All I commit to is the idea that the concept doesn't exist as a part of our communal means of making sense of the world.

Do all tokens count as reproduction? This is difficult to answer. It brings up heavily contested questions of concept identity. On my account, conceptual reproduction occurs when an individual token is reproduced, and this supports the continuation of a particular classificatory practice. Perhaps a better question is: how much can a token deviate from ordinary practice before it stops being an instance of conceptual reproduction? After all, at some point the copying mechanism will be so deviant that it will not help sustain the continuation of a conceptual practice. I am not sure what degree of deviance is permitted. Suppose that someone talks about human flesh as food, when we would ordinarily not classify humans as food. Does this count as conceptual reproduction? Perhaps this is a borderline case. It wouldn't be odd for someone to respond, 'Humans aren't food?' Yet, it also wouldn't be odd for someone to say, 'Humans are food, we just don't eat them.' I am doubtful that there is something common to all uses of a term.[92] To be clear, as explained in chapter one, I take concept identity not to be a matter of sameness of knowledge structure between any two people within a community. Rather, concept identity involves being a part of a particular representational tradition. Thus, conceptual reproduction, on my way of thinking, tolerates some deviance from ordinary usage. What is required is that one enters into *de facto* or automatic coordination via natural language signalling strategies. This constitutes a kind of *family resemblance* view.[93] A token must conform to some aspect of ordinary practice to count as conceptual reproduction. Consequently, this allows for deviance to some degree. I leave open what these amounts are.

The permissibility of deviance allows for a diversity of conceptual practices to emerge and propagate within a community, which can disrupt existing patterns of coordinated thought and talk. Importantly, a conceptual engineer can '[drop] a mutation into a population, hoping that it will 'reproduce' and in one way or another establish itself, and even replace all of its alternatives over time' (Richard 2020, p. 377). In other words, conceptual reproduction is not completely conservative. While it does function to maintain the continuation of an existing conceptual practice, its imprecision plays a role in shaping the practice in different directions. And this is something that we, as activists and engineers, are able to take advantage of.

Given the foregoing, we might define conceptual reproduction roughly as follows:

> **Conceptual reproduction** occurs in a thinking and speaking community $L$ when a word-token associated with a classificatory disposition $W$ is non-accidentally copied from prior stable patterns of classification with $W$ in $L$.

On this account, there are different ways that tokens copy prior usage. In a sense, conceptual reproduction is almost trivial. Any conceptual practice, that isn't radically deviant, counts as reproduction. Calling vegetables 'food', with the intention to pick out a stable kind food, and roughly conforming to the classificatory practices of a linguistic community, counts as reproducing the coordinated concept FOOD. My goal is to pick out important cases. Specifically, I will identify cases of conceptual reproduction that perpetuate unjust social relations, which raises normative questions regarding how we might stop the force of precedent moving a conceptual practice forward. Plausibly, it is the reproduction of such conceptual practices that must be discontinued in order to successfully implement competing justice-promoting concepts.

---

[92] That is, I am doubtful that there is a sharp distinction between competent and incompetent uses of a term. There will be vague borderline cases.

[93] This relational view contrasts with Fregean accounts in which there is a particular pattern of usage that constitutes conceptual similarity.

### 3.3. Local Reproduction and Global Patterns of Reproduction

It's important to note that conceptual reproduction takes place at different socio-ontological levels, with varying degrees of formality.

*Local* reproduction occurs when a classificatory disposition is reproduced within a restricted local context; a one-off instance of reproduction within a thinking and speaking community. Broadly, this can be divided into two kinds: individual and collective. Individual local reproduction occurs when a single person tokens a word that copies prior stable patterns of classification, such as in conversation. This can be conscious or otherwise. A speaker might deliberately choose to use a word for a particular purpose (e.g., efficient communication or to display their identity),[94] or they may have acquired dispositions that conform to existing patterns of usage through unconscious reinforcement mechanisms (e.g., social pressure). Collective local reproduction occurs when a concept is reproduced by a group agent, such as governments, and the World Health Organization, etc. For instance, a government might use a particular concept, such as WOMAN, when drafting a bill about maternity leave. The key point of local reproduction, whether it is individual or collective, is that it captures just one use of a term at a given time, in a way that copies prior usage, rather than a pattern of uses across contexts.[95]

*Global patterns of reproduction* is a result of a large number of individuals, within a linguistic community, consistently and reliably engaging in forms of local (individual and collective) reproduction.[96] That is, global patterns of reproduction emerge from many instances of local reproduction that produce a robust regularity of individual and collective classification within a community. 'Global' patterns of reproduction aren't restricted to entire communities, consisting of many different social sub-groups. It simply refers to the general tendencies of members of a particular community to classify one way rather than another, which includes groups such cricket clubs to NASA. Reproduction at this level involves stable individual conformity and coordination of thinking and speaking practices, and this helps achieve mutual understanding within a community. To be clear, I am interested in the global patterns of reproduction of entire communities consisting of many different social sub-groups; the reproduction of concepts that constitute, at least in part, a public language. Concepts that are reproduced on this scale are reasonably common, and the sorts of practices that are up for social philosophical enquiry, such as the common or ordinary concepts WOMAN, MAN, FOOD, RACE, etc.[97]

To be clear, the relationship between local reproduction and global patterns of reproduction is as follows: when there are robust regularities in local reproduction within a community, or consistent and reliable classificatory practices that result from copying word-tokens, then this constitutes a global pattern of reproduction. And, importantly, there are many different mechanisms or social

[94] Alex Davies (2020) has argued that in a conversation in which there are alternative and competing ways of precisifying a context-sensitive term or expression, choosing one precisification over another can be a matter of displaying an aspect of one's identity.

[95] Though, this is not to say that the speaker who uses the word at a time isn't disposed to use the word in all contexts. Instead, local reproduction simply captures a time-stamped usage; whether or not that person is inclined to use the word in the same way across different speech situations.

[96] This might constitute standard public meaning, but I will refrain from committing myself to that position here.

[97] Of course, there can be similar concepts of say, woman, that only operate in particular sub-communities. For example, a wholly trans* inclusive concept of woman might not be produced on a global scale, but is reproduced within progressive circles.

forces at work that keep local reproduction going, pushing a classificatory practice with a term forward, and sustaining the global patterns of reproduction. Such social forces can be seen pushing the reproduction of concepts at different levels of formality.

### 3.4. Formal Reproduction

Formal (or institutional) reproduction occurs in established and recognised social institutions. In such institutions, there are decision-making procedures and structural processes that affect the people for whom the institutions were designed. Typically, in such institutions concepts, or forms of classification, are embedded in laws and rules, policed by formal institutional authority rather than (just) informal authority of social pressure.

Consider the example above. Suppose a government passes a piece of legislation about how many weeks of maternity leave a woman can take (with no similar provision for fathers or other caretakers), which was informed by certain stereotypes surrounding women being natural caretakers. In a patriarchal society, this does a number of things. It reproduces a patriarchal coordinated concept of women which is partly responsible for organizing nuclear families in a way that unfairly distributes gender roles – it builds into the legal framework the norm that (only) women should be supported as the initial caregivers for infants. The woman stays at home and tends to domestic life; the man takes care of business. This translates into the unjust denial of certain life opportunities for women (e.g., opportunities to gain wealth, status, and power). In this case, a community-wide classificatory disposition is reproduced insofar as 'woman' is used in a way that conforms to prior stable patterns of usage.[98]

This is a reasonably clear example of a conceptual reproductive mechanism. It explicitly involves formal, or institutional, local reproduction that helps sustain global patterns of classification with a term. However, consider an example in which conceptual reproduction is implicit. Suppose that a government passes a law that bans the live export of farmed non-human animals. Of course, there are a number of concepts that are reproduced explicitly, such as NON-HUMAN ANIMAL. But think of what is *implicitly* reproduced. When a government passes such a law, there is a certain thought-process behind this: animals are *food*, but changes need to be made to the course of moving bodies from factory farms to the dinner plate. In this case, what we can see is that the concept FOOD is involved in formal local reproduction, sustaining global patterns of reproduction, even though it is never explicitly expressed or deployed in the piece of legislation in question. Moreover, this utilizes and reinforces certain characterizations, in particular stereotypes, such as cows being a healthy or desirable food-option, or that they are a part of a Western diet, etc., which also contributes to pushing the classificatory practice forward.

To sum up, formal reproduction, within institutions, takes place within structured groups, with decision-making processes and policing procedures, that use words in a way that copies prior stable patterns of classification. As discussed, parliament is one place where this occurs. But there are other locations. Groups such as local councils, crickets clubs, schools, and the like, can reproduce concepts in that domain, while at the same time contributing to global patterns of reproduction.

---

[98] Embedding the coordinated concept of woman in the legal system scaffolds and encourages reproduction insofar as the legal system enables and constrains what one is able to do. So for example, a trans* person will have to reveal the sex they were assigned at birth in cases where official documents do not give the option for trans* people to state their gender identity. And because the government is seen as an overarching *epistemic* authority, this motivates people to think and speak in the same gender-exclusive way.

### 3.5. Informal Reproduction

Like formal reproduction, informal reproduction involves the use of a term that copies prior stable patterns of classification. How it differs is that informal reproduction occurs in less structured ways. Informal reproduction is structured in some sense, involving certain processes and procedures, but it tends to operate below the level of awareness, often being a matter of unconscious bias or (willful) ignorance. Moreover, the concepts that are reproduced are not embedded in law, nor backed by institutional authority.[99] Such reproductive mechanisms feature in our everyday existence. They are operative in any normal conversation or ordinary interaction with the world. Because of this, many cases of informal reproduction are quite mundane. However, there are many other cases to which we should be attentive. I have in mind informal reproductive mechanisms that involve illegitimate power dynamics; the oppressive social forces that push dominant practices of classifying forward, and which help maintain systems of unjust dominance and subordination.

Informal reproduction takes place at different ontological levels. At one level, informal reproduction occurs as an interplay between social sub-groups within a broader linguistic community; the reproduction of intercommunally coordinated concepts via intergroup social dynamics. Informal reproduction of this kind can be seen in Miranda Fricker's (2007) work on *hermeneutical injustice*. To reiterate, hermeneutical injustice occurs when there is a gap in the collective hermeneutical resource; the set of concepts that are shared across sub-groups. This gap exists due to structural injustice, or what Fricker calls *hermeneutical marginalization*: structural identity prejudice that prevents marginalized group members from participating in the production and dissemination of concepts shared within a broader linguistic community. While Fricker brings attention to a specific epistemic wrongdoing, what she also does, perhaps unintentionally, is highlight how concepts are involved in a series of informal local reproductions that contributes to sustaining global patterns of classification. Because marginalized groups are (often) excluded from the production and dissemination of concepts, dominant groups, who are typically unmotivated to change, will continue to engage in certain practices with words that copy prior stable patterns of classification – it serves them well, after all. In other words, dominant concepts rarely come up against alternatives to compete with that stand a genuine chance of replacing them. Consider an example.

Imagine a time before we had the concept of sexual harassment. Much of the behaviour that we now count as sexual harassment might have simply been passed off as flirting, and therefore harmless. That is, there was a gap in the collective hermeneutical resource that meant we missed out on making proper sense of the unjust treatment that many women experienced. Part of the reason why this gap existed was because dominant group members, such as men, constantly reproduced FLIRTING in contexts that were better understood as involving harassment. This prevented alternative understandings, developed from marginalized situatedness, from making their way into public consciousness. After all, the introduction of competing forms of classifying the same behaviour produces a tension. Given our general characterizations about flirting and harassing, this blocks a framing that sees flirting and harassing occurring together. And because our classificatory dispositions are undergirded by characterizations, it seems that we cannot classify behaviour as both *harmless* and *harassment* at the same time. Thus, the higher social standing of one group makes their classification of particular behaviour easier to reinforce, and harder to question. This counts as informal reproduction insofar as 'flirting' was regularly and consistently tokened in local contexts in a

[99] Religious sensibilities are a difficult case to explain.

way that copied prior and dominant classificatory practices,[100] therefore sustaining regularities in classification within a community (i.e., global patterns of reproduction).[101] What's important to note is that while there was resistance from dominant groups to take up this concept, we now exist in a world in which SEXUAL HARASSMENT is a part of our shared set of coordinated concepts. So, it is possible to overcome pernicious hermeneutical gaps that are maintained by factors such as structures that reward ignorance and misogynistic behaviour. In saying this, however, there are still concerted efforts by dominantly situated people, such as cis men, to contest the proper application of the term 'sexual harassment' to contexts or events that are most appropriately understood as sexual harassment (i.e., gaslighting). I will talk about this idea, namely gaslighting at the level of conceptual exchanges, in chapter seven.

### 3.6. Social Forces and the Conceptual Status Quo

Throughout the discussion so far, I have spoken about different of kinds social infrastructure responsible for pushing conceptual reproduction forward. I want to draw out some of these mechanisms in this section. I will make explicit social forces that work on local contexts to causally bringing about local reproduction, and which sustain regularities in classification within a community.

One might say that reproducing a concept simply owes to the stability of a classificatory practice. That is, we unconsciously copy word-tokens because it is an entrenched practice that we often do not question. However, this story is incomplete. There are social forces and infrastructure that are responsible for keeping a classificatory practice stable, which contributes to causing individuals, consciously or not, to produce a particular word-token. And the list of such forces and infrastructure, at different levels of formality, is extensive – too extensive to list them all. To be clear, it includes, among other things, social pressure and hierarchies, norm compliance, institutional policing and authority, social policing and authority, imitation of trendsetters, revered artwork and literature, common architecture, and even things like Twitter, Facebook, and YouTube algorithms, echo chambers and epistemic bubbles, communicative structures such as media presses, the education system, high-profile speaking platforms, testimonial chains, common standards for assessing evidence, 'official' historical narratives, monuments, etc.

Only some of these will be unique to formal reproduction (e.g., certain kinds of institutional requirements or norms); others will be seen at any level of formality (e.g., social pressure). To be more concrete about ways that social forces keep things conceptually as they are, a clear example of a strong mechanism or incentive for copying prior stable patterns of classification is legal punishment and benefits which require paying attention, or accepting, certain legal categories (e.g., *marriage*, *illegal immigrant*). For example, a racist society bent on ensuring that foreigners don't steal their jobs

[100] Dispositions that maintain a rape culture which obscures or masks certain wrongs perpetrated by men

[101] Though marginalized groups often are *prevented* from contributing to the set of coordinated concepts shared within a broader linguistic community, there are times at which groups do engage in a kind of negotiation as to which concepts will continue to survive and organize our lives, and which concepts get the axe. This negotiation is not conducted through formal proceedings with structured processes. Instead, it occurs simply when groups engage in a kind of 'conversation'. Deborah Mühlebach (2019) calls this a 'semantic contestation'. Mühlebach shows, using the race term 'black', that the different discursive practices of different communities come into conflict with another, each vying to be the dominant understanding in the general linguistic community. Groups contest which conceptual practices should be stabilized in order to determine how our lives ought to be organized. When illegitimate power gets involved, however, dominant groups are often able to have their preferred concept receive widespread uptake.

will be strongly motivated to push practices with term 'undocumented migrant' forward in order to subject refugees to the full force of the law.

I don't have time to go through each of these social forces, and how they work to push local conceptual reproduction forward, and sustain global patterns of reproduction. Nevertheless, what I will say is that conceptual engineers, motivated by political aims, should be interested in researching or examining each of these mechanisms of conceptual reproduction and how they contribute to maintaining the conceptual status quo. Only with such understanding will we get the full picture of just how difficult the implementation challenge can be. Below, and for the remainder of this thesis (chapters four, five, six, seven, and eight), I will focus on one such mechanism. Namely, the social forces that encourage local reproduction in interpersonal speech situations, such as conversation. My interest in this mechanism of reproduction owes, at least in part, to the fact theorists working on conceptual implementation have made suggestions about how to bring about conceptual innovation through reproducing alternative concepts, rather than dominant ones, in conversational contexts – though, not in the language of reproduction. I will argue that being able to disrupt the local reproduction of dominant concepts in conversation is a tall order, and likely implausible, given the social forces operating in such contexts that work to push the reproduction of dominant concepts forward, and to maintain the conceptual status quo.

What I have touched on, also, is that conceptual reproduction is often guided and reinforced by *power relations*. This is clear in the case of formal or institutional reproduction in which the state has authoritative power to pressure individuals into coordinating their thinking and speaking practices with a term (e.g., through enforceable laws). When this pressure is effective, our beliefs will reflect the expectation that users conform to such practices. Thus, institutional norms can reinforce overlapping patterns of mutual expectations in the community. This isn't merely to do with enforcing subordinating *social* practices. Rather, it is about enforcing certain dispositions with words, which implicitly propagates broader usage of such words within a linguistic community. In other words, institutions often play a crucial role in sustained local reproduction of concepts, and thus, global patterns of reproduction.

Perhaps less obviously, power dynamics also infect local informal reproduction, such as conversation between individual speakers. As mentioned, conceptual reproduction in this form appears trivial – we do it all the time. We informally reproduced concepts simply by using words in a way that conforms to prior stable patterns of classification. However, like inter-group dynamics, illegitimate power also operates within interpersonal contexts such as conversation, playing a role in locally reproducing existing conceptual practices, and sustaining global patterns of classification. Some have gestured to this idea. Kristie Dotson (2012) introduced the notion of *contributory injustice* to explain a distinctive epistemic wrong that occurs when dominant group members fail to take seriously the concepts developed within marginalized situatedness. In other words, the concepts developed from within the margins are willfully ignored by dominantly situated people.[102] Think of a woman who deploys a concept of sexual harassment in order to get across to her boss that he is acting in a morally problematic way. But, imagine, upon hearing this, the boss simply says, 'Oh, that was just harmless flirting.' In this case, the dominant concept FLIRTING is informally and locally reproduced owing to the fact that the boss, in willful ignorance, failed to recognise and classify his behaviour as

[102] Gaile Pohlhaus Jr.'s (2012) notion of *willful hermeneutical ignorance* is similar. For Pohlhaus, active ignorance, causally connected to social structure, is responsible for why dominant group members refuse to give uptake to concepts developed from within marginalized situatedness, therefore further contributing to hermeneutical marginalization.

sexual harassment.[103] And this individual-level interaction connects up with, and helps sustain, broader patterns of usage with the term – patterns that maintain hierarchical systems of injustice.

Another theorist who brings attention to the power at play in conversation over concepts is Jonathan Jenkins Ichikawa (2020). He introduces the notion of *contextual injustice* to explain situations in which disadvantageous standards for knowledge, and other context-sensitive expressions, are introduced into conversation. This occurs through the raising of pernicious error possibilities. For example, when assessing testimony describing sexual harassment, one might employ unnecessarily high standards for knowledge by saying something like, 'Are you sure it wasn't an accident?' This tends to play on existing stereotypes that frame our thinking about the relationship between 'woman' and 'knowledge.' That is, women, by and large, are thought to be irrational and less likely to have knowledge. By raising the standards, the man's question reflects and reinforces this pattern of understanding and underlies dominant classificatory dispositions within the community.

Dotson and Ichikawa, it can be argued, illuminate a specific way that concepts are reproduced: dominant speakers are able to exercise power in conversation in order to have their preferred concept 'accepted' as apt or appropriate for the context.[104] More specifically, and what will be explored in the remaining chapters, dominant speakers rely on their greater social authority to gain accommodation of the conceptual understanding in a conversational context. And this is typically through contrasting alternative classifications of a subject, event, or behaviour. It's important to recognise the implications of this. Because coordinated concepts play a substantive role in affecting the individualistic content of our thoughts, when such concepts are 'accepted' in conversation this frames what we perceive as possibilities for action – it triggers certain action-scripts, or a practical consciousness, that undergirds behaviour. Consider an example. Suppose two friends, George and Kelly, are getting ready to go to a party and they are arguing over the status of a jacket.

George: The jacket is black, that doesn't count as smart-casual!
Kelly: But it doesn't have shoulder pads?[105]

We can see that George and Kelly are arguing over which concept should be expressed by 'smart-casual' in the context. In more detail, we might say that each is inviting the other to 'try on' a particular characterization, which, if accepted, will play a substantive role in underlying or explaining their classification of the jacket as either smart-casual or not (e.g., George and Kelly disagree as to whether the property *black* or *having shoulder pads* should be prominent and central in their overall thinking about the jacket).

Suppose that George and Kelly have the aim of achieving consensus. There are practical consequences for the concept that is accepted. In particular, it will determine whether Kelly will wear the jacket. In other words, the accepted concept, or how the jacket is classified in virtue of its associated characterization, plays a role in providing an action-script that will structure Kelly's

[103] Dominant concepts fail to compete with alternatives developed from marginalized situatedness under conditions that would put each concept on an equal footing. That is, marginalized concepts, or those concepts develop within marginalized communities, aren't able to be compared to dominant concepts in contexts that would show just how much better the marginalized concept is. Thus, we might think that marginalized concepts need to be able to flourish in 'safe space,' or spaces in which they can develop before they enter into the mainstream. This 'space' is perhaps something akin to Charlie Crerar's (2016) notion of an 'expressively free environment.'

[104] I use quotation marks because in conversation with illegitimate power dynamics, acceptance of a concept is often coerced – such as gaslighting.

[105] This is an instance of *metalinguistic negotiation* (Plunkett and Sundell 2013).

behaviour in response to a resource (i.e., a jacket). One might think that such cases aren't all that important. Perhaps this is true. However, some cases are very troubling. To go back to the example above, think of a situation in which a woman is sexually harassed by her boss, and the boss then gaslights the woman into thinking that such an act doesn't count as 'sexual harassment' by triggering certain stereotypes about women being notoriously irrational. Or think of a White person who argues with a Person of Colour that merely mentioning that 'n' word doesn't count as 'racism' by triggering certain stereotypes about marginalized people being 'too woke.' In such cases, dominant speakers have the backing of whole societies, with norms and histories in their favour, which increases the chances of having their preferred concept 'accepted' as apt or appropriate for the context, and therefore locally reproduced.

What's the take away? Informal local reproduction occurs in socially significant situations (even if they don't *seem* significant), such as when speakers resolve a dispute as to whether an event counts as sexual harassment or an instance of racism. This is because oppressive extant coordinated concepts are often endorsed by speakers who have greater control over the joint activity of pairing concepts with words in a context. Concepts are informally and locally reproduced in such cases owing to the operation of illegitimate power dynamics that tend to favour existing patterns of classification. And, further, the struggle over which concept is apt or appropriate for a context can produce injustice within that context, such as silencing the subordinate perspective. Social infrastructure of this kind, such as prevailing norms of conceptual engagement in interpersonal contexts, serves to reinforce the conceptual status quo within the broader community. Exactly how will be spelled out in the remaining chapters of this thesis.

## 3.7. Characterizations and Conceptual Reproduction

The foregoing has mostly focused on the reproduction of classificatory dispositions. But one might wonder: how are characterizations related to conceptual reproduction?[106] It's important to note that social forces also work to maintain dominant characterizations within a community, such as the psychological convergence mechanisms discussed in chapter one – including institutional power and authority, social norms and conventions, epistemic authority and exposure, etc. And these forces often work with the logical structure of characterizations to resist revision (i.e., prominence and centrality). Let's explore this idea.

One might expect that we continuously update our characterizations in light of discrepant information, or counterexamples that do not conform to elements of a characterization.[107] For example, one might update their stereotype that hipsters care about blends of coffee given exposure to a number of hipsters who think that coffee is disgusting. But gradual inductive change only seems possible in cases where individuals are responsive to inductive evidence, such as accepting that a particular counterexample poses a risk to the elements of one's characterization. Problem is, individuals often protect their ways of seeing by rejecting or ignoring inductive evidence, especially when such individuals have a strong investment in a particular characterization. For example, it seems

[106] I should make it clear that I accept that characterizations can be reproduced, at least to a significant extent. And this raises an interesting question about the mechanisms that stabilize practices of copying characterizations within a community. In chapter one, I discussed one kind of mechanism: psychological convergence mechanisms. However, there will be other social forces that maintain dominant characterizations within a community. Unfortunately, I do not have time to discuss that here.

[107] In empirical work on stereotype change, this is called 'the bookkeeping model' (Bicchieri and McNally 2018, p. 35 see also Crocker et. al., 1984).

that there is resistance to overturning the pernicious stereotype that Muslims are threats to Western society (e.g., terrorists) given the perceived (yet irrational) cost of failing to 'protect' oneself and their country.

Or, instead of gradual inductive change, one might expect that in light of enough discrepant information, counterexamples that target the *core* elements of a characterization (i.e., prominence and centrality), that we will completely revise our intuitive understanding of a subject matter.[108] However, whether one recognises or accepts that a particular piece of information counts as a counterexample will be subject to different social norms, such as whether one takes a particular news outlet or research institute to be 'credible' given their political views (e.g. Trump supporters who reject CNN); as well as cognitive defence mechanisms, such as the self-enhancing qualities, including self-esteem, that one derives through in-group identification (Kahan et. al. 2007).

Or, even if discrepant information does not dislodge an existing characterization, or elements of a characterization, we might expect that counterexamples are hived off as an exception to the rule, and a new, but associated, characterization is developed.[109] Again, this will depend on whether the counterexample is accepted or recognised. However, another issue is that this 'hiving off' can maintain pernicious ways of thinking by protecting the core of a pernicious characterization. For instance, if one has racist attitudes about Asian people being bad drivers, and yet one comes across an Asian person whose driving ability is greater than average, one can simply develop a new stereotype that can deal with this seeming contradiction. This appears to be a means of preserving ideological ways of seeing. And it seems that what might be required to dislodge a pernicious characterization is social encouragement or enforcement that motivates people to completely revised their stereotypes rather than subtype them.

There is much more to say about the relationship between characterizations and social forces, which, for the most part, is an empirical matter that goes well beyond the scope of this thesis. However, enough has been said to show how it is that the social forces that maintain dominant characterizations, such as the psychological convergence mechanisms that stabilize similarity of associative thought within a community, relates to conceptual reproduction. One of the key roles of a characterization is their use in higher order cognition. In particular, as discussed in chapter one, characterizations underlie or explain our classificatory dispositions; our intuitive understanding of a subject matter that tells us what is significant or important about a social category, and this motivates us to classify one way rather than another (e.g., if we take self-identification to be significant to gender, we will classify individuals as a particular gender on the basis of this). Given this relationship, when overlapping characterizations are maintained within a community, via mechanisms that secure psychological similarity, so too are community-wide classificatory dispositions. Put differently, the social mechanisms that contribute to stabilizing overlapping characterizations between members of a thinking and speaking community *also* help stabilize dispositions to classify aspects of the world with a particular term.

Importantly, the stability of dominant characterizations helps push current practices of classification forward via social processes of copying insofar as it constrains our recognition of alternative classification possibilities. If all we see as important about our gender categories is secondary sex characteristics, or something of the like, then this will severely limit our understanding of new and better ways of classifying that emphasises self-identification. We often reproduce

[108] This is called 'the conversion model' (Bicchieri and McNally 2018, p. 35 see also Rothbart 1981; and Rumelhart and Norman 1978)

[109] And this is called 'the subtyping model' (Bicchieri and McNally 2018, p. 35 see also Rothbart 1981).

dominant classificatory practices just in virtue of the fact that prevailing characterizations close us off to new and better ways of seeing and speaking. Thus, in order to disrupt mechanisms of conceptual reproduction, one area of social life that we might target are the mechanism that maintains similarity of associative thought within a community. Having better characterizations is apt to bring about better means of classifying aspects of the social world as belong to the same kind.

### 3.8. Implementation and the Ideology of Individualism

With the idea that social forces entrench existing classificatory practices (and dominant characterizations) on the table, I want to revisit the objection I mentioned in passing in §3.1. To reiterate, I claimed that the implementation challenge, while lacking a proper formulation, is often posed too individualistically. This is clear in the formulations of the implementation challenge that were discussed at the beginning of this chapter. For example, Herman Cappelen and David Plunkett claim that:

> Once you have settled on an ameliorative strategy, you might want to do some work to implement it, that is, you might want to engage in a bit of activism on behalf of your ameliorative strategy. If that's something you want to do, it raises an 'implementation challenge': how are ameliorative strategies best implemented? (2020, p. 3).[110]

And, this appears to be echoed by Alexis Burgess and David Plunkett:

> There is what we can call attempts at conceptual implementation. Roughly, this involves a kind of advocacy, in which one tries to get some people (ranging from a lone individual to a large population), to actually take up and use the concepts one is in favor of (2020, p. 5).[111]

Of course, there is not enough said in these passages to infer a fully-fledged theory of conceptual implementation. However, each passage says enough to raise some red flags. Importantly, both formulations of the implementation challenge focus mostly on strategies of individual conceptual advocacy or activism. Conceptual change appears to be a matter of what an individual can do to convince others in their community to adopt an alternative conceptual practice. The issue with this framing is that it incorrectly shapes our perception of the shape, size, and nature of the problem. If all that can be done to bring about change to a conceptual repertoire is for an individual to simply persuade others, or undertake advocacy on the part of their preferred concept, then of course it will be a confronting, daunting, and seemingly insurmountable exercise to pitch an ameliorative strategy to a world that has entrenched ways of seeing and classifying bolstered by the social forces, such as mechanisms of conceptual reproduction, that maintain the conceptual status quo – especially those forces that preserve naturalised understanding of socially contingent facts, giving the appearance that certain aspects of the world are unchangeable or inalienable.

---

[110] Mark Richard calls this an *A-project*: 'a project of offering and trying to get others to accept a revisionary ameliorative analysis of a concept…' (2020, p. 360).

[111] Other theorists have put forward suggestions as to how one should think about the implementation problem, but nevertheless bring in assumptions about the nature of the problem, such as focusing on meanings within a metasemantic externalist framework (e.g., Jorem forthcoming). I use these quotes because they are reasonably neutral with respect to what a proposed alternative concept might look like.

With this said, individualistic accounts of implementation are understandably irresistible. Appealing to thoroughgoing social processes and structures tend not to appease our cravings to centre the autonomy of people in changing the world – we want heroes to revere, and villains to condemn. Nevertheless, when it comes to the conceptual engineering of social kind concepts, we shouldn't forget its political roots. The goal of such projects, including my own, is to advance pursuits in the project of social justice. And this has never been a lone-wolf endeavour. Thus, we need to avoid the 'ideology of individualism' (Haslanger 2015, p. 9). This is the idea that while we should recognise the role of individual agency in the explanation and disruption of unjust social structures, we should not focus on this alone. Instead, we need to further acknowledge and emphasise that 'social life consists of individuals embedded within complex social relations governed by culture and both constrained and enabled by resources made available through those very social relations' (ibid. 2015, p. 9, see also Tilly 2002, p. 6). It is important we appreciate the relationality of our social lives, and that in virtue of this relationality we are enabled and constrained with respect to what we can achieve – entrenched structural systems, the social infrastructure that maintains the conceptual status quo, can be a serious barrier to change. In the context of conceptual engineering, this means that we must resist centering or emphasising individualistic implementation strategies.[112] When it comes to the implementation challenge, we should look outside the individual. Individuals, given their relational situatedness, are limited in what they are able to do in order to persuade others in a community to take up an ameliorative strategy. There are many social forces, especially, but not limited to, mechanisms of conceptual reproduction, that get in the way of successful individual advocacy of an ameliorative strategy. And such forces tend to favour dominant terms of conceptual engagement.

Another reason to resist individualistic implementation strategies, or individual conceptual advocacy, is that many of the concepts that we intend to bring into our collective set of coordinated concepts, in order to bring about more just social relations, have been developed within non-dominant communities. That is, improving our existing conceptual repertoire often involves transplanting a conceptual practice from one community into another in the hope that it will spread far and wide. For example, many concepts, such as SEXUAL HARASSMENT (Brownmiller 1990), were developed in consciousness raising groups, consisting of many people who contributed to shaping the concept and getting it to the point where it was able to flourish within non-dominant contexts. From this, such concepts were able to take root and propagate in dominant contexts, becoming a part of our shared coordinated concepts. However, this inclusion does was not a result of isolated individuals engaging in activism on the part of their concept. Rather, it was a matter of shifting an entire community of practice into a new domain of use. In other words, it took many people, over a long period of time, to develop a concept and to bring it into broader public consciousness. I will say more about this in the concluding chapter.

Given this, one might wonder whether the problems with individualistic implementation strategies dissolve when we think in terms of groups or communities, rather than just individuals. A theorist who endorses a view of this kind is Steffen Koch (forthcoming). He argues, presuming a moderate semantic externalism, that while it is impossible, or at the very least implausible, that individuals have 'immediate control' over reference, groups nevertheless have 'long-range control.' That is, individuals do not have the ability to perform one, uninterrupted action that will have an intended consequence for a community's conceptual repertoire. However, *collectives* can bring about such

[112] It's important to note that Haslanger is not talking about the implementation challenge here. Rather, her argument concerns theoretical *explanations* generally.

intended consequences through (potentially) uninterrupted actions performed over an extended period of time (Koch forthcoming, p. 12).

Moving from individuals to groups or communities is certainly better. It offers a more promising option when thinking about how we might implement an ameliorative strategy. However, this shift doesn't free us entirely from the ideology of individualism. [113] Rather, it changes the relevant 'individual,' or agent, that is responsible for bringing about conceptual change. Looking back at the quote from Haslanger, what we need to do in order to avoid the ideology of individualism is to recognise the fact that individuals are embedded in *social relations* that are both enabling and constraining; to recognise that we are situated in structures that work to maintain the conceptual status quo. Shifting the analysis of implementation to the level of groups does not automatically achieve this recognition. Groups, despite having more power than individuals alone, are still enabled or constrained in virtue of their social positionality. In particular, groups, like individuals, will be subject to mechanisms and incentives of conceptual reproduction that work to push a conceptual practice forward, such as institutional authority, social hierarchies, in-group dynamics, communication structures (e.g., media, schools), etc. Given this, group conceptual advocacy, or group-level implementation strategies, still face many barriers to effective conceptual innovation. There is a great deal of social infrastructure that prevents individuals and groups from encouraging others within a community to change their minds; especially to a point where a sufficient number of individual minds can change the global conceptual practices within a community. Thus, in order to develop tailored-made implementation strategies that stand a chance of being successful, we must work out exactly what it is that individuals and groups can do; and despite groups being able to do more than individuals, there are still limitations on what groups are able to achieve when it comes to conceptual change. Paying attention to the features of our social structure that maintains the conceptual status quo, such as mechanisms of conceptual reproduction, allows us to see what our limitation are, and where we might be able to make conceptual interventions. What we will see is that not all groups have equal power, and for some the job is much harder. In particular, we will see that the implementation challenge is thoroughly shaped by power relations. I will have much more to say about this in chapters five, six, seven, and eight.

Let me sum up the thoughts of this section. (Re)engineered concepts, in particular justice-promoting concepts, must work with or against the social forces that are (partly) responsible for maintaining dominant terms of conceptual engagement if they are to be implemented in the conceptual repertoire of a community of thinkers and speakers. Specifically, such concepts must be reproduced if they are to gain community-wide uptake. The problem that faces individualistic implementation strategies, or individual conceptual advocacy, is that it cannot be understood as means of effectively exploiting or disrupting conceptual reproductive mechanisms; the social forces that push a conceptual practice forward. Put differently, individualistic implementation strategies will struggle to compete with mechanisms of conceptual reproduction, part of our social infrastructure, that tends to favour dominant terms of conceptual engagement. Thus, we must come up with other implementation strategies, beyond individual conceptual advocacy, that will be able to deal with the structures that keep dominant concepts in place.

[113] Despite this, Koch's project is not political – another draw back of his theory. While he gives us a framework in which we might understand how to achieve reference change as a community, it does not frame the challenge of implementation in a way that makes explicit the forms of power that we might confront in our efforts to revise oppressive coordinated concepts. Moreover, Koch's account only focuses on reference change, with an externalist metasemantic framework. He does not think about changing knowledge structures.

## 3.9. Conclusion

Conceptual reproduction occurs all the time. And what's more is that conceptual reproduction, while allowing for a certain degree of variation in classification from ordinary usage, is mostly conservative. That is, conceptual reproduction tends to favour dominant patterns of classification. What makes this a problem is that implementing an ameliorative proposal into a community that has established and entrenched classificatory dispositions must negotiate, and potentially overcome, mechanisms of conceptual reproduction in order to be successful. This makes the implementation challenge rather immense. We must resist conceptual reproduction, a conservative mechanism that stabilizes classificatory dispositions within a community, in everyday conversation, engagement with material reality, inter- and intra-group dynamics, and in formal processes, etc. How to go about such resistance varies. Different contexts call for different tactics.

An upshot of recognising these different conceptual reproductive mechanisms is that we can either work with or against them. It seems that we can exploit formal proceedings in order to bring about conceptual change, or we can undermine them. We can use different means of disrupting the illegitimate power that operates between groups or individuals, or we can bypass such power altogether. In some sense, we need to engage in conceptual *de*-production. This involves disrupting the history of particular patterns of usage in order to undermine actual dispositions, associations, and social roles. However, to disrupt this history we need to find ways of dealing with existing mechanisms of conceptual reproduction. And it seems that in virtue of the fact that conceptual reproductive mechanisms are conservative that individuals, in their effort to bring about conceptual change, will struggle to compete with systems that naturally tend to maintain the conceptual status quo. Thus, individualistic implementation strategies appear to be an inadequate means of dealing with or overcoming the structures that regulate our dispositions to classify one way rather than another.[114]

For the most part, I have spoken abstractly about the social forces (partly) responsible for maintaining the conceptual status quo and which make the success of individualistic implementation strategies unlikely. My aim for the rest of the thesis is to speak more concretely by casting doubt on the plausibility of a *particular* individualistic implementation strategy that has gained popularity in recent conceptual engineering literature. This strategy involves an attempt to exploit or disrupt mechanisms of conceptual reproduction (and psychological convergence) that operate within conversational contexts.

Focusing on conversation as a location of conceptual resistance is understandable. After all, we communicate with others everyday of our lives. As such, conversation is a domain in which classificatory practices and characterizations are regulated through processes such as social pressure, norms of communication, desire to be understood, encouragement and shaming, etc. So, if we are to stand a chance at disrupting existing patterns of classification, it seems that conversation, especially with dominant group members, is an important target. In particular, and what many theorists seem to

[114] A relevant analogy is climate change and normative theorizing. Achieving a substantive reduction in atmospheric carbon dioxide and methane levels is not something that any individual can do on their own. And, given that individuals cannot bear obligations over those actions that they are unable to perform, then placing a normative burden on individuals to solve climate change is inappropriate (Collins 2019). Though implementation individualism is not a thesis about determining who has an obligation to implement an ameliorative proposal (cf. Podosky 2018), the same reasoning can be applied. Individuals have little chance of overcoming the mechanisms that regulate classificatory dispositions and similarity in characterization within a community, so perhaps individuals should not be the focus of implementation strategies.

be primarily interested in, we should be concerned with *metalinguistic disagreement* (e.g., Plunkett and Sundell 2013; Cantalamessa 2019; Sterken 2020; Davies 2020). Before I can explain why metalinguistic disagreement is an unlikely location at which the implementation of an ameliorative proposal will be successful, I must first spend time in the next chapter laying out some theoretical machinery that will help us to understand what metalinguistic disagreement is, and why it is thought to be an important kind of speech situation for the purposes of conceptual change. Specifically, I must introduce a particular mechanism of conceptual reproduction that I call *conceptual accommodation.*

# Chapter Four

## Conceptual Accommodation

*Speech acts, including directives generally, and hate speech specifically, can acquire authority by an everyday piece of social magic: authority gets presupposed, and hearers let it go through, following a rule of accommodation.*

Rae Langton 2018, p. 152

### 4. Introduction

In recent literature on conceptual implementation, there has been a focus on what speakers can do within conversation to convince a hearer to adopt an ameliorated conceptual practice. Specifically, theorists have been concerned with what speakers can do within speech situations in which there is disagreement *about* concepts (i.e., metalinguistic disagreement), or at the very least, a chance for conversation about which concepts should be used in the context (e.g., Cantalamessa 2019; Sterken 2020). The aim of the speaker is to plant the seeds of a new conceptual practice in a local context, typically with a dominantly situated hearer, in the hope that it will then be passed along to others and propagate far and wide within a community (e.g. Pinder forthcoming, p. 17).

My aim in this chapter is to explain that this individualistic implementation strategy, that is, the strategy of getting a hearer to adopt an alternative conceptual practice in a local context via metalinguistic disagreement, often involves having to compete with strong social forces that work to maintain the conceptual status quo, and which suppresses conceptual change. Put differently, there is a particular conceptual reproductive mechanism within conversation about concepts, and which influences its kinematics, which reduces the likelihood of successful individual conceptual advocacy of a non-dominant conceptual practice. There are structural barriers within conversation that tend to favour dominant patterns of classification.

In saying this, reiterating what was said at the end of the previous chapter, the significance of this domain, as a site for conceptual engineering, is understandable. Conversation is a part of our everyday practices. Because of this, it is a location that we should be concerned to correct if we are to ensure that ideological concepts are not being constantly reproduced. And, to stress this even further, much of our conduct within speaking situations is automatic. We needn't consciously reflect, at least under normal conditions, on permissible or standard means of exchanging words. It comes naturally to us, most of the time. Because of this, unjust power often lurks behind conversational norms or conventions that we take for granted; norms and conventions that must be overcome in order for alternative conceptual practices, or ameliorative strategies, to be implemented. However, such norms and conventions are not likely to be overcome in the context in which they are operative. And, subsequently, making interventions within conversation will often fail at undermining or resisting the reproduction of ideological concepts.

One might wonder: if I am so interested in social structure, why focus on the conceptual reproductive mechanisms that occur in local contexts? There is already much attention given to the

kind of reproductive mechanisms that involve intergroup dynamics, and how such mechanisms can be corrected so as not to reproduce ideological concepts – though, this discussion is not in the language of 'reproduction.' Miranda Fricker (2007) has argued that in virtue of the structural identity prejudice, marginalized groups are often unable to participate equally in the production and dissemination of the cognitive (and linguistic) resources that are shared across social sub-groups within a community.[115] This is a distinctive epistemic wrongdoing that she calls *hermeneutical injustice.*[116] I have previously spoken of the details of hermeneutical injustice, so I won't reiterate them here. All that I want to draw attention to is the idea that hermeneutical injustice involves the reproduction and reinforcement of existing concepts, including patterns of classification and characterizations, by dominant groups at the level of social structure. It is a means by which social structural procedures, through an interplay between social sub-groups, affect the available cognitive resources that individuals are able draw on and utilise in their interactions with others and material reality.

I want to move away from speaking about such processes of conceptual reproduction. Instead, I will focus on the kinds of reproductive mechanisms that occur at the local level. I am interested in the means by which individual decisions and behaviour causally affect the concepts shared by members of a thinking and speaking community. To put it differently, rather than explaining how it is that *intergroup* dynamics, or lack thereof, play a causal role in partly determining the concepts shared by different social sub-groups, or those who share a public language, I will explain how it is that *interpersonal* dynamics, between individual people, maintain and reinforce such concepts.

It's important to note that there is existing literature that identifies reproductive mechanisms of this kind – again, not in the language of 'reproduction.' For example, Kristie Dotson (2012) has argued that, through willful ignorance, dominant speakers fail to take up the cognitive resources developed from within marginalized situatedness when presented to them in situations like conversation. A similar idea is defended by Gaile Pohlhaus Jr. (2012). I will not focus on willful ignorance to explain the reproduction of coordinated concepts in conversational settings. Though I acknowledge that it is often, and perhaps even likely, present. Instead, I will explore the power-infected kinematics that shape *metalinguistic* discourse – disagreement or conversation about the concept that is most apt or appropriate for a context relative to certain aims and goals. In particular, I will examine the notion of *conceptual accommodation.*

Lewis's (1979) notion of *accommodation* has received much attention in recent socio-political theorising. Much of this interest stems from a concern for how speech acts are involved in the creation of oppressive conversational contexts (e.g., Langton and West 1999; McGowan 2009, 2012, 2018; Popa-Wyatt and Wyatt 2017; Langton 2018).[117] This discussion is mostly attentive to complex linguistic phenomena such as the accommodation of *presupposition* and *illocutionary force.* I argue that we can gain further understanding of the role that accommodation plays in the creation of oppressive conversational contexts by examining how it is that our more basic ways of representing things become a part of conversational common ground. In other words, we should understand how our attitudes about *concepts* are accommodated in conversation. This has hitherto been undiscussed,

[115] See Podosky (2018) for how conceptual resources are defined.

[116] Ishani Maitra (2017) argues that hermeneutical injustice might be best construed as a wrong that is distinctively linguistic.

[117] There are many more theorists interested in oppressive speech (not all of whom are interested in accommodation). See MacKinnon (1987, 1993), Langton (1993, 1998, 2015, 2018), Butler (1998), McGowan (2003), Maitra (2012), Kukla (2014), and McKinney (2016). This chapter is a contribution to this broader discussion.

or at least taken for granted. Yet its significance in understanding how concepts are reproduced, often under conditions of injustice, and where we might locate sites for ideological resistance, makes it worthy of emphasis and exploration.

In this chapter, I will restrict my discussion to cases of conceptual accommodation that occur within *neutral* contexts. That is, those conversations in which conceptual accommodation occurs, but illegitimate power is *not* (assumed to be) operative. I recognize that this is idealistic. And I want to move away from ideal philosophy of language.[118] After all, conceptual accommodation has the potential to sustain the existence of conceptual practices that can structure our lives in oppressive ways. I will discuss power-affected cases of conceptual accommodation in chapters five, six, seven, and eight. Before I get to this, however, I must first provide a definition of conceptual accommodation. This requires introducing a taxonomy.

## 4.1. Score and Ground

Like cricket, conversation proceeds under rules that govern the conduct of participants. Such rules determine the *score* of conversation. Given this, conversational score is an abstract and normative structure. It is partly descriptive insofar as it registers moves made by players in a language game; recording the conversational participants' shared set of information. We might think of this as the background conditions for conversation (Kölbel 2011, p. 51). But, score is primarily normative in the sense that it *constrains* the contributions of participants. This is through the specification of correct play. Speakers abide by a directive to operate within certain rules, such as the rule only to offer contributions that are relevant to the question under discussion (i.e., Grice's cooperative maxims). In other words, conversational moves are tracked in score, and the state of play, in conjunction with the rules that govern conduct, constrain the behaviour of speakers by determining the permissibility of futures moves.

Score differs from *common ground*. Each tracks a different aspect of conversation. While there are competing accounts of common ground, I take inspiration from Rae Langton (2018). According to Langton, common ground consists of the psychological states of players; it is the set of attitudes that speakers share at a time in conversation. When a move is made in a language game not only does score change but so does common ground. Score changes first, and a change in common ground is a causal consequence. (Langton 2018, p. 154).[119]

## 4.2. Accommodation

Despite keeping track of different phenomena, 'common ground functions in much the same way as conversational score' (McGowan 2018, p. 192). Score and ground place normative requirements on

[118] Cappelen and Dever (forthcoming) contest whether the ideal/non-ideal distinction can be made in the philosophy of language. They argue instead that there are simply 'underexplored' topics. Of course, this is true. But one might wonder *why* such topics have been underexplored – have power relations corrupted our interest in discovering uncomfortable truths about the world? Moreover, one wonders whether Cappelen and Dever do not seem to explore whether *conceptual engineering* can be either ideal or non-ideal, which is surprising given that it is a through-going form of normative enquiry. This is my interest.

[119] Though, there might be occasions in which score has changed but there has been no change to common ground. For instance, a speaker might say something that isn't registered by the hearer. While the score is updated to include what the speaker has said, there isn't a shared attitude in common ground about what was said given that it wasn't heard by at least one party.

speakers. Permissible conversational moves depend on: (i) the previous moves that have been made in conversation, *and* (ii) the shared psychological states of players. When something is said by a speaker that is *impermissible* given the state of play, score and ground can be updated following a *rule of accommodation*.

Accommodation occurs when a hearer adjusts the context of a speech situation to include information that renders a speaker's apparent norm violating utterance correct play in order to advance conversation on a subject matter (Adams 2019, p. 3). In other words, accommodation involves the following: (i) A speaker attempts to make a contribution to conversation, but the information that determines the state of play, and the rules that govern the conduct of participants, is such that the contribution is in violation of permissible conversational moves, but despite this, (ii) a hearer accepts the contribution as meeting the necessary requirements to count as correct play. To see this, consider *presupposition* accommodation.

Presupposition accommodation occurs when a speaker makes an assertion that presupposes a proposition that is not part of the speakers' shared set of information, and a hearer accepts[120] the presupposed proposition, rendering the speaker's utterance correct play (Lewis 1979). For instance, imagine that Fil and Fig are eating lunch, and Fig says 'it's my daughter's birthday tomorrow.' Prior to this, Fil did not believe that Fig had a daughter. There was no shared attitude toward the proposition *Fig has a daughter*. Thus, in order for Fig's utterance to count as correct play, and to have its intended effect on conversation, Fil first needs to revise his belief about Fig and her status as a mother. Fil needs to update common ground following a rule of presupposition accommodation.

Notice that presupposition accommodation concerns the accommodation of presupposed *propositions.* And, some think that common ground only includes shared attitudes, typically beliefs, toward propositions (Stalnaker 2014). However, it seems that not all psychological states that can be shared are propositional attitudes, and such other states play an important role in advancing conversation on a topic. For example, Rae Langton talks about the accommodation of affective attitudes or emotions (2012, p. 140). Thus, common ground consists of more than propositional attitudes. Given this, we can think of presupposition accommodation as a *general* species of accommodation involving the updating of common ground that amounts to the following: A speaker makes a move in conversation that requires an attitude to be shared with a hearer, but the required attitude is not shared, so the hearer updates common ground by adjusting their psychological state to match the speaker's, rendering the move correct play. Following Langton, call this *perlocutionary accommodation* (2018, p. 155).[121] Specific to my interests, common ground also consists of a set of certain attitudes toward *concepts.*

### 4.3. Conceptual Attitudes

My interest is in social kind concepts. In particular, I am interested in the relationship between such concepts and accommodation. To see this relationship, it's important to note that we often hold attitudes about the *aptness* or *appropriateness* of a concept for a context – it is a judgement or assessment about whether a certain classification applies correctly to an individual, behaviour, or

[120] 'Acceptance' is a broad term for any positive propositional attitude (Stalnaker 2014, p. 4).

[121] In contrast to *illocutionary accommodation*, which occurs when a speaker attempts to perform an illocutionary act that requires certain felicity conditions be met, and the hearer changes the context such that the felicity conditions are met (e.g., Langton 2014, p. 4; Adams 2019).

event.[122] We might call these *conceptual attitudes.* Such attitudes are distinct from those toward propositions insofar as they are attitudes toward concepts. Attitudes toward concepts are not attitudes toward things that can be true or false.[123]

For example, imagine that Fil and Fig are arguing over whether Fig is cool. Fig says, 'Of course I'm cool; I only wear black.' Fil responds, 'That's not what cool is; wearing black is so 2015.' Here, the disagreement is not at the object-level. It concerns the attitude one should have toward Fig's concept of cool given their situatedness in a particular socio-historical milieu. Fig believes that the concept is appropriate because it invokes a characterization of a certain colour (or lack thereof) that is typical of prevailing fashion. Fil thinks that this attitude is dated. While this argument concerns what counts as cool, it also concerns how Fil and Fig should structure their overall thinking about 'coolness'

And, to reiterate, this is not merely about the things that *count* as 'cool.' It also concerns how Fil and Fig should structure their overall thinking about 'coolness.' Why? Characterizations are in complex interaction with classificatory dispositions. That is, characterizations tell us what is important or significant about a category or kind. Because of this, characterizations underlie or explain our higher order cognitive competencies, such as classification. So while the argument is about what counts as cool, how this argument can be resolved is by convincing the other to take on the relevant perspective that will prompt or motivate a revised classificatory disposition.

One might argue that Fil and Fig are in fact having a first order disagreement over the truth-value of the proposition, *Fig is cool.* If this is the case, then it appears that the relevant attitude *is* propositional, not conceptual. However, such a disagreement would only make sense in the context of a shared attitude about the aptness or appropriateness of a particular concept. For the disagreement to be first order, Fil and Fig must agree, to some extent, on answers to questions such as: Which concept of cool should we be using? What features should be central and prominent in our understanding of coolness? In other words, Fil and Fig must settle on a concept to be expressed by 'cool' in the context in order to determine whether COOL appropriately or correctly applies to Fig, or whether Fig is fitting of the characterization of coolness.

Given that common ground consists of the attitudes that speakers share, should we say that it also includes *conceptual* attitudes? I think the answer is reasonably clear: yes.[124] After all, concepts are important for the flow of conversation. The conceptual attitudes in common ground at a given time determines the range of utterances that can be successfully interpreted by participants, the subject matters of conversation that can be spoken on, and the moves that speakers are able to make. For instance, if a hearer lacks a concept, and therefore a conceptual attitude, required to make sense of a

[122] I will remain intentionally vague about what makes a concept apt or appropriate for a context. One reason is because aptness or appropriateness is sensitive to one's aims and goals. For example, a concept of spicy is apt or appropriate relative to whom a dish is being served to (i.e., practical reasons); or the concept of whale is apt or appropriate relative to the aims and goals of scientific inquiry (i.e., epistemic reasons). Another reason is that my primary interest isn't in 'objective' or neutral standards by which we assess the aptness or appropriateness of a concept, but instead how it is that certain concepts are forcibly or coercively *rendered* apt or appropriate for a context – typically through unjust power relations.

[123] Remember, a stereotype can be apt without the features that make up the stereotype being true of the people to whom the stereotype is applied (e.g., hipsters needn't be interested in blends of coffee or have beards to be hipsters).

[124] Moreover, Langton herself says, 'I want to propose... the idea that the phenomenon of accommodation might extend beyond belief – beyond conversational score, and common ground, as originally conceived – to include the accommodation of other attitudes...' (2012, p. 140). Langton has in mind attitudes such as *desire* and *hatred.* However, I see no reason why conceptual attitudes wouldn't count as being a candidate for the sort of thing that can be accommodated, and thus feature, in common ground.

speaker's utterance, then the hearer is unable to interpret, at least not completely, what the speaker has said. Failure to grasp what was said means that certain subject matters of conversation are off the table, and hence limits the range of permissible moves that speakers can make in that conversation. Further, if a hearer rejects that the concept expressed by a speaker is apt or appropriate for the context, then, despite being able to correctly interpret the speaker's utterance, there are implications for conversation: the contested concept becomes the topic of discussion, instead of the things to which the concept is supposed to apply. For example, disputants might initially begin conversation by arguing over whether *S* is tall, but then transition into discussion as to which concept should be expressed by 'tall' in the context (i.e., determining criteria for 'tallness'). The conversation moves from first order discussion about the world, to second order disagreement over the aptness or appropriateness of a conceptual practice. Again, this limits the range of moves that participants are permitted to make. I will bring these points out in more detail in §5.7 and §5.8.

### 4.4. Conceptual Accommodation

We're at the point where we might begin to see the mechanics of conceptual accommodation. Roughly, it involves the accommodation of a conceptual attitude in a context. But what situations require that one accommodate a conceptual attitude? To explain this, I need to introduce one final notion: *conceptual moves*. A conceptual move is a speech act in a language game that involves the use of a concept, or a disposition to use a word in a certain way. That is, when it comes to concepts, a conceptual move simply involves the deployment of a concept as expressed by a particular (non-indexical) term. [125]

With all of the relevant taxonomical machinery on the table, I am now in a position to provide a rough definition of conceptual accommodation. In its most general form, conceptual accommodation amounts to the following:

> **Conceptual accommodation** occurs when a conceptual attitude is accommodated into common ground, which has the causal effect of rendering conceptual moves involving the attitude correct play.

In other words, conceptual accommodation involves the updating of common ground to include an attitude that a particular concept is apt or appropriate for the context. This partly determines the range of conceptual moves that can be registered in conversational score.

There are different ways that conceptual accommodation can occur. And how a conceptual attitude is accommodated into common ground depends on a particular kind of *conceptual interaction*. Broadly, a conceptual interaction is an event in a language game in which a speaker says something that expresses a concept C (i.e., a conceptual move), and the speaker presumes that the

[125] There are also moves that attempt to get a hearer to take on a characterization or a way of seeing the world. In a sense, such moves area kind of *proposal.* That is, with the use of a term or expression, the speaker proposes to a hearer to frame their overall thinking in a way that matches their own. For example, when Fil says to Fig, '...wearing black is so 2015' he is inviting Fig to (re)structure her thoughts in a way that makes the attributed feature of being 'out-dated' prominent and central. Why a proposal? We cannot force a person to see the world in the same way we do. Instead, through certain utterances, we can bring to salience particular features that will apt to cause a hearer to adjust their thinking. The notion of proposal that I've offered appears to be neutral in the sense that, on the face of it, the speaker is simply putting forward something that a hearer is permitted to reject. Of course, this is true some of the time. However, proposals are not always friendly. They can be put forward through coercive tactics or implicit psychological manipulation (i.e., gaslighting).

hearer possesses[126] C in order to make sense of what was said. And by 'make sense,' I mean something along the Gricean lines of 'correctly identifying the communicative intention of a speaker.'

Given this, conceptual interactions can either be *complete* or *incomplete.* A complete conceptual interaction is thus:

> **Complete conceptual interaction:** An event in a language game in which a speaker says something that expresses *C* such that a hearer requires possession of *C* in order to make sense of what the speaker has said *and* the hearer possesses *C* (and thus can make sense of what the speaker has said).[127]

Put differently, a speaker utters something in conversation that expresses a concept and the hearer possesses the concept that enables them to successfully make sense of the speaker's claim involving the concept.[128] Compare this to an incomplete conceptual interaction:

> **Incomplete conceptual interaction:** An event in a language game in which a speaker says something that contains *W* that expresses *C* such that a hearer requires possession of *C* in order to make sense of the speaker's use of *W and* the hearer *does not* possess *C* (and is unable to make sense of the speaker's use of *W*).

Here, the speaker's utterance cannot be entirely understood since the hearer does not possess the relevant concept needed to make sense of the speaker's claim involving the concept.[129] For example, in a tirade about how horrible it is that chickens are debeaked in factory farms, all of the moves that I make with respect to this will have very little intended effect on score since much of what I say cannot be interpreted by a hearer in virtue of lacking possession of the concept DEBEAKING.[130] And, in virtue of lacking this concept, the person subject to my tirade will be unable to fully structure their thoughts in a way that matches mine. The hearer is unable to take on the relevant conceptual attitudes that would enable us to advance conversation on the subject matter of the treatment of chickens in factory farms.

This distinction between kinds of conceptual interaction allows us to identify general forms of conceptual accommodation, distinguished by their respective temporal directedness in conversation. Notably, there are three: (1) permitting something as correct play (*Permission*); (2) establishing what play can be made (*Establishment*); and (3) negotiating what play has been made (*Negotiation*).

---

[126] Possession comes in degrees, for example see Burge on Frege (1990), Bealer (1998), Higginbotham (1998). Here, possession implies some degree of competence: recognizing when someone is deploying a concept in conversation. For more on possession conditions, see Burge (1979), Peacocke (1992), and Schroeter (2008).

[127] Complete conceptual interactions aren't so demanding as to require *full* possession of concepts. And given that possession comes in degree, the distinction between complete and incomplete conceptual interactions exists along a spectrum. That is, there will be cases where an individual only has partial possession of a concept, and thus it cannot be said that the conceptual interaction is strictly complete or incomplete. However, explanation of the degree of possession required for a complete conceptual interaction goes beyond the scope of this thesis.

[128] Sometimes the concept that needs interpreting may occur at the level of what has been presupposed or implicated. I will only focus on the concepts that have been expressed.

[129] We might accept that the hearer has a very basic grasp of the concept, but not enough to understand the speaker's claim involving the concept.

[130] A hearer might not understand a speaker's utterance for many reasons. Here, I am just interested in those cases in which the hearer does not possess a concept.

*Permission* concerns what it is happening at a present time in conversation; *Establishment* concerns what can happen at a future time in conversation; *Negotiation* concerns how to interpret what has happened at a past time in conversation. Furthermore, *Permission* and *Negotiation* are complete conceptual interactions, and *Establishment* is incomplete. Finally, *Establishment* and *Negotiation* are forms of conceptual accommodation.[131]

## 4.5. Permission

*Permission* is perhaps the most well known form of accommodation. Lewis's *rule of accommodation for presupposition* provides the clearest picture of how this works in conversation. Lewis argues that if a speaker utters something whereby it depends on a presupposition being shared by conversational participants, and which the hearer has not presupposed, then if the hearer does not object to the presupposition, the presupposition enters into score (Lewis 1979, p. 340). In other words, the hearer *permits* the presupposition to enter into score.[132] Presupposition accommodation is a complete conceptual interaction. A hearer will permit a speaker's presupposition when, minimally, there is a shared conceptual attitude(s) in common ground that enables the hearer to understand or grasp what is being presupposed.

I bring up *Permission* only to contrast it with forms of conceptual accommodation. In cases of *Permission,* accommodation operates within the context of mutually understood and accepted utterances.[133] A speaker says something, and immediately the hearer knows how to interpret what was said, and the hearer does not question **(**publicly) whether what was said is apt or appropriate for the context. From this, we can ask two questions that go beyond Lewis's notion of presupposition accommodation:

(i) What happens in conversation when a hearer does not know how to make sense of a speaker's utterance owing to a conceptual deficiency?

(ii) What happens in conversation when a hearer disagrees with a speaker as to whether a concept is appropriate for the context?

To (i), there are (at least) four possible answers. First, the hearer makes a mistake about what the speaker is trying to convey by using the wrong concept (or characterization) to interpret the utterance. Second, the hearer ignores the speaker's utterance either because they realize that they have failed to understand what was said, or else because they believe that the speaker has made a mistake. Third, the hearer possesses the relevant concept needed to make sense of a speaker's utterance, but the speaker has not conveyed enough information to allow the hearer to discern which concept has been expressed, or how a hearer should structure their thoughts. Fourth, the speaker's use of a concept that the hearer does not possess means that the hearer updates their set of

[131] *Permission* is a form of illocutionary accommodation; *Establishment* and *Negotiation* are forms of perlocutionary accommodation.

[132] Presupposition accommodation is not the only form of *Permission*. See Lewis (1979).

[133] Understanding and acceptance are pre-conditions for illocutionary accommodation. For perlocutionary accommodation, understanding and acceptance are not pre-conditions. This is because it concerns changes to the attitudes of speakers, and gaining understanding or accepting a meaning constitutes change to the attitudes of speakers.

interpretive tools to make sense of the utterance.[134] I am interested in the last of these options. I am concerned with cases in which a speaker deploys a concept in conversation that is not part of the hearer's individual conceptual repertoire, but the hearer comes to acquire the concept, and subsequently takes on the relevant conceptual attitude. We might call this *the rule of accommodation for interpretation.* When this rule is observed, it *establishes* the moves that can be made in conversation. The hearer is able to make sense of conceptual moves in virtue of acquiring a concept and taking on the attitude that the concept is apt or appropriate for the context.

To (ii), we might expect that conversational participants will engage in a *disagreement* as to which concept is the most apt or appropriate for the context. In such cases, speakers disagree over what should be expressed by a shared term, or how a term should be used (i.e., how to classify with a term).[135] When speakers aim at consensus in such disagreement, in order to achieve a practical end, this is called *metalinguistic negotiation* (Plunkett and Sundell 2013, Davies 2020).[136] Settling disagreement of this kind is important. Not only for practical reasons outside of conversation, but for the sake of conversation itself. For participants to advance conversation on a topic, they will need to accept the aptness or appropriateness of a particular concept, or take on the relevant conceptual attitude, in order to determine the moves that can be made in that conversation.

Neither of the suggested answers to (i) and (ii) constitutes an instance of *Permission.* Further, both are instances of conceptual accommodation. Thus, I will spend time explaining the mechanics of two forms of conceptual accommodation: *Establishment* and *Negotiation.*

### 4.6. Establishment

*Establishment* is an incomplete conceptual interaction, unlike *Permission.* Participants to conversation *do not* share a particular concept, and therefore lack a shared conceptual attitude, at a given time in a language game that would enable a hearer to fully interpret a speaker's utterance. In other words, a speaker attempts to update score by performing a conceptual move, but the hearer cannot entirely make sense of what was said owing to a conceptual deficiency. For example, a speaker might say, 'Philosophical inquiry into gender relations should be intersectional,' and because the hearer lacks possession of INTERSECTIONALITY, and possibly cannot structure their thinking in a way that understands social categories as mutually constitutive,[137] the hearer cannot fully grasp what the speaker has said.

---

[134] This can happen explicitly (e.g., the hearer asking, 'what do you mean by that?'), or tacitly (e.g., the hearer putting together bits and pieces of information to acquire the concept being expressed).

[135] And this dispute might come down to which attributed features should be made contextually prominent and central in their joint understanding.

[136] Note that this isn't strictly the definition of metalinguistic disagreement provided by Plunkett and Sundell (2013). Despite this, the goal of achieving consensus in metalinguistic disagreement, or a disagreement over which concept to be expressed by a shared word, is implicit in Plunkett and Sundell's discussion of a fictional disagreement between Oscar and Callie over the content that should be expressed by 'spicy':

> Why would Oscar and Callie consider it worth their time to engage in such a disagreement, when they already agree on what the chili actually tastes like? ... [B]ecause how we use words matters. For Oscar and Callie, as for many of us, an agreement amongst all the cooks in the kitchen that the chili can be described as 'spicy' plays an important role in... decision-making about whether to add more spice (ibid. 2013, p. 15).

[137] I leave it open as to whether it is possible that one can fail to possess a concept yet still structure their overall thinking in a way that matches a speaker who is deploying that concept. It seems plausible that this could happen. For example, before women had the concept of sexual harassment to make sense of particular

In cases like this, a speaker's utterance has little intended effect on score in virtue of an absence of a shared conceptual attitude in common ground. And, the hearer does not have the required conceptual attitude because they do not possess the concept needed to make sense of the use of a word in an utterance. Hence, in order for score to be updated in a way that the speaker intends, the hearer needs to acquire the relevant concept, and take on the attitude that such a concept is apt or appropriate for the context. We might say that the speaker needs to make the conceptual interaction complete.

When a hearer acquires the concept needed to make sense of a speaker's use of word in an utterance, and takes on the required conceptual attitude, the participants are *establishing* the moves that can be made in conversation. Namely, they are making it permissible to perform certain conceptual moves that will update score in a way that the speakers intend. Continuing our example, when a hearer updates their conceptual repertoire to include INTERSECTIONALITY, and takes on the conceptual attitude that such a concept is apt or appropriate for the context, they are making it possible to perform speech acts about intersectional analyses (that will be understood).

What's important to note about cases of *Establishment* is that the acquisition of a concept means that certain psychological associations, or characterizations, often come along for the ride. That is, we cannot help but organize our thoughts in a certain way when we learn a new concept. Elisabeth Camp (2017) argues that this is for three reasons.

In the context of conversation, particularly with respect to metaphorical usage, Camp explains that conveying a characterization has certain effects on a hearer. First, when a speaker utters something that expresses a characterization, it produces *complicity* in the hearer since simply making sense of what has been said requires that the hearer 'mold [their] mind in the speaker's image: to structure his overall thinking so that the relevant features really are intuitively prominent and central for him, in a way that goes significantly beyond the hypothetical contemplation of a proposition' (Camp 2017, p. 51). Second, taking on a characterization can be *irresistible*. This is because characterizing is a disposition to structure one's thoughts. Hence, when a speaker offers a characterization, or a way of seeing, in conversation, a hearer will, more or less automatically, interpret a situation to match the speaker's thoughts (Camp 2017, p. 49). Third, one cannot straightforwardly *deny* a characterization when it has been offered in conversation. This is either because the hearer has already interpreted a situation in the way that a speaker intends, or because 'repudiating any one feature leaves the overarching organizational structure, in its intuitive application, untouched' (Camp 2017, p. 52). Moreover, these dispositions are supported by the fact that members of the same linguistic community are typically subject to the same psychological convergence mechanisms that regulate or stabilize similarity of associative thought.

For example, suppose that Raffee and Carlos are discussing police brutality against Black Americans. During conversation Raffee says, 'Systemic racism is a real problem, you have to be woke,' to which Carlos responds, 'What does 'woke' mean?' Suppose Raffee then goes on to explain that it is about being aware or conscious of the surreptitious ways that White supremacy can manifest in supposedly protective institutions. Upon hearing this, Carlos acquires a concept of woke, and thus comes to understand what Raffee has said. This, however, has certain psychological effects on Carlos.

---

wrongdoings, it seems that some women, nevertheless, were able to structure their overall thinking in a way that interpreted particular situations as wrong in the same way as other women who had similar experiences. Moreover, this seem required to make sense of how the concept of sexual harassment came about – namely, through consciousness raising groups in which members exchange stories, recognized similarities in experience and thought, and developed a concept to capture such similarities (see Susan Brownmiller (1990) for the historical development of the concept of sexual harassment).

He cannot help but structure his overall thinking in a way that matches Raffee in virtue of coming to understand what Raffee meant by the use of 'woke.' That is, Carlos develops a way of seeing, with a concept of woke in mind, that includes, *inter alia,* packaged characterizations of American policing (e.g., violent, unjust) and its oppressive relationship to Black Americans (e.g., subjugated, persecuted).

In this case, the psychological effect on Carlos appears to be a good thing. After all, it is important that people are aware of the systemic and brutal racism against Black Americans at the hands of the police. However, not all cases of *Establishment* will involve positive or even merely 'neutral' psychological effects. Some involve harmful characterizations that can render contexts oppressive. This will be explored in chapter eight.

## 4.7. Negotiation

*Negotiation* is a complete conceptual interaction, like *Permission.* Participants to conversation share a concept at a given time in a language game that enables a hearer to make sense of a speaker's utterance. However, unlike *Permission*, *Negotiation* involves a situation in which an utterance is made that expresses a concept but conversational participants disagree as to whether the concept correctly applies to the context. Speakers do not share the relevant conceptual attitude because which concept should be used in the context is contested.

To see this, consider an example offered by Plunkett and Sundell:

> 'Suppose that Oscar and Callie are cooking together... Oscar and Callie have each just tasted the chili, and Oscar utters [(a)] while Callie utters [(b)].
>
> (a) That chili is spicy!
> (b) No, it's not spicy at all
>
> In this case, it is much less natural to think that there is some antecedently settled, objective fact of the matter about the contextually salient threshold for "spiciness". Rather than advancing competing factual claims about some independently determined threshold, it seems more natural to think of Oscar and Callie as negotiating what that threshold shall be (2013, p. 14 – 15).'

As Plunkett and Sundell suggest, it is tempting to see this exchange as a back-and-forth over whether the chili is, in fact, spicy. However, instead of interpreting the dispute as concerning whether some independently determined criteria of spiciness has been met, we can think of it as the interlocutors negotiating which criteria is best for the context. They are arguing over which concept of spicy, out of a competing range, should be used – or how to classify the object in question. Plausibly, this disagreement is not simply just a matter of how to classify a subject matter. Rather, given the complex interaction between characterization and classification, it is a disagreement that concerns which features of the dish in question should be most salient in one's overall thinking, such as bean-to-spice ratio, the relevant group that the dish is being served to, or how the dish compares to hotter alternatives, etc. I will have much more to say about this in the next chapter. For now, suffice it to say that each speaker is making a conceptual move in the hope that a certain conceptual attitude will be

accommodated into common ground. Importantly, the speakers are aiming at consensus in order to achieve a practical end (Plunkett and Sundell 2013, p. 20).[138]

Settling disagreements of this kind is important for conversation. If interlocutors have not come to an agreement on the concept that a word will express in a context, or a shared frame of mind, then there is very little means of furthering conversation on the subject matter at play. To see this, let's continue with our example. In the conversation, Oscar offers an utterance that is challenged by Callie – there is disagreement over which concept of spicy is most apt or appropriate to apply to the chili. We might ask: What happened to Oscar's first utterance? Did it have an effect on conversation? It did. Though, conversation didn't evolve with respect to the participants' exchange over the subject matter of chili.[139] Instead, it shifted to the metalinguistic level, and the conversation moved to the subject matter of spiciness – i.e., which concept of spicy is apt or appropriate for the context? Imagine how the conversation might have proceeded:

Oscar: Of course it's spicy! We're cooking for Australians, remember?
Callie: Yeah, but I know these guys. They love spice!

The topic of conversation has shifted: it has moved from a discussion about chili, to a discussion about spiciness thresholds, or how to classify the chili. And in the disagreement, competing characterizations are introduced, where certain features are proposed as being made more central than others: Oscar brings up that the kind of people who will eat the food are Australians, Callie says that these particular Australians enjoy spicy food. What we can see is that chili moves to the background of conversation, and spiciness takes front and center.

Why is this important? The conversational topic shifted so that the speakers can settle on which concept of spicy is most appropriate for the context. And coming to an agreement on a concept of spicy, through getting another to re-structure their thoughts about the spiciness of the dish, allows conversation on the subject matter of chili to advance in a particular direction. That is, speakers can make moves that concern the chili once it is settled which conceptual attitudes will be common ground. Imagine the conversation once this has happened:

Oscar: Fine, the chili isn't spicy.
Callie: Excellent; let's add more spice.

In this final exchange, two things have happened. First, Oscar has changed the effect he intends to make on score by updating common ground (i.e., accepting that a particular concept of spicy is apt or appropriate for the context). Thus, instead of including Oscar's move that the chili is spicy, it is now registered as including Oscar's move that the chili is not spicy. Second, because Oscar and Callie have settled on a concept, and updated common ground to include a shared conceptual attitude, they

[138] Whether a concept is chosen is a subjective, pragmatic matter of participants imposing or bargaining for their preferred concept. This resembles *ad hoc* concepts in Relevance Theory (Carston 1997; Sperber and Wilson 1998): 'categories made on an occasion for the needs of that occasion' (Allott and Textor 2012, p. 187).

[139] Perhaps with the exception that the score includes something like, *it is impermissible to talk about the chili being spicy until it is decided what 'spicy' means*.

are able to make moves that advance conversation about how to practically-orient themselves in response to the chili. Each can register a move in score about adding more spice.[140]

As we can see, conceptual accommodation has occurred. Oscar has accommodated Callie's conceptual attitude(s). This was a practical necessity for conversational purposes. If Oscar or Callie didn't agree on a concept to be expressed by 'spicy,' then it would have been difficult to have a substantive conversation on the subject matter of chili. The next chapter will be a closer examination of the mechanisms of metalinguistic negotiation that lead to conceptual accommodation.

## 4.8. Conceptual Accommodation and Conceptual Reproduction

How does the accommodation of conceptual attitudes relate to conceptual reproduction? To reiterate, conceptual reproduction occurs when a word-token associated with a classificatory disposition is non-accidentally copied from prior stable patterns of classification within a community. Importantly for this discussion, I will focus on local and informal conceptual reproduction. Recall from the previous chapter that local reproduction occurs at a local scale, such as when a concept is reproduced in conversation between friends; and conceptual reproduction is informal when it occurs outside of established institutions (e.g., at a party). Thus, in the event that an accommodated conceptual attitude involves the use of a word that copies prior stable patterns of classification, this counts as an instance of informal local reproduction. And, given the conservative nature of conceptual reproduction, this helps sustain the existence of the conceptual practice, moving the force of precedent forward, causing further tokenings within the community. Consider an example.

Let's say that the concept of spicy that Callie endorses is the dominant concept; a classificatory disposition stably associated with a particular word that tends to be copied within a community – say it's a community that loves spicy food. When Oscar accommodates Callie's conceptual attitude, Oscar does not just update common ground in the context. Rather, he contributes to reproducing the concept on a local scale. In other words, by accommodating the conceptual attitude, Oscar also plays a part in the local reproduction of the dominant concept of spicy. Moreover, because Oscar accommodates a conceptual attitude of a concept that is used far and wide within a community, Oscar's local reproduction contributes to the global reproduction of that concept. And further still, given the relationship between concepts and social reality, Oscar's accommodation helps stabilize coordinated behavioural dispositions around what we eat, or how we eat.

One might think: Why does any of this matter? Sure, the accommodation of a conceptual attitude contributes to conceptual reproduction, but this seems rather innocuous. Whether a dominant concept of spicy is reproduced is of little consequence, or maybe only of consequence for those who don't like spice. This is somewhat true. In some communities, it won't matter whether a concept of spicy is reproduced. But, in other communities, those rife with unjust power relations supported by ideological concepts, conceptual reproduction through the accommodation of conceptual attitudes can be a real problem, even for concepts like SPICY and its associated characterizations. Why? Food is a political matter. What is considered tasty, or disgusting, is caught up in ideology that presupposes insider-outsider relations. For example, a predominately White community might look upon the food of South Asian or South-East Asian people as disgusting, inedible, or otherwise inferior in virtue of the fact that it is not 'Western hot' but 'Asian hot.' So, if a conceptual attitude is accommodated that consists of a concept of spicy that alienates the cuisine of 'outsiders,' this is a deeply political act.

[140] It's possible that Oscar and Callie won't settle on the *same* concept of spicy, but will accept different concepts of spicy, both of which will apply to the chili.

We can move beyond the example of spicy to see why conceptual accommodation matters when it comes to the reproduction of concepts. Here is one example. Suppose that a Person of Colour is speaking to a White person at a party in a predominately White neighbourhood. The White person asks where the Person of Colour is from, and the Person of Colour says, 'I'm Filipino,' to which the White person says, 'Don't you guys eat dogs?' Now, imagine that the Person of Colour says, 'that's racist,' and the White person contests this by saying, 'Come on, that's not racist, it's true.' The speakers are now engaged in metalinguistic disagreement. They are disputing which concept should be expressed by the word 'racist' in the context. Given that the dispute is taking place in a White neighbourhood, the White person is speaking with a whole history of White supremacy or racism behind them, as well as social norms that mask this fact. Due to his discomfort, and pressure to conform, let's say that the Person of Colour accommodates the conceptual attitude of the White person; he accepts that the concept, and the characterization that underlies this concept, endorsed by the White person is apt or appropriate for the context. In virtue of this, the Person of Colour plays a role in the local reproduction of an ideological concept of racism that oppresses him directly. His act of accommodation is a contribution to a larger system of global reproduction of an ideological concept, which subsequently helps entrench pernicious social and epistemic structures – namely, organizing our epistemic and practical lives in a way that upholds and masks systems of unjust hierarchy. This will be explored in more detail in the remaining chapters.

## 4.9. Conclusion

The foregoing case represents what I will call *non-ideal metalinguistic disagreement.* And this kind of disagreement will be the subject matter of the next chapter. Before I move onto discuss non-ideal disagreement, I will sum up what has been said.

In this chapter, I introduced particular conceptual reproductive mechanisms that occur within conversation; specifically, within metalinguistic discourse. I identified two types of conceptual accommodation: *Establishment* and *Negotiation*. The former occurs when a concept is needed to be acquired in order to advance conversation on a topic; the latter occurs with the resolution of a disagreement as to which concept should be expressed by a word in a context. Because conversation is a part of our everyday lives, such mechanisms are crucial to the maintenance of ideological concepts. After all, metalinguistic discourse is a routine conversational process of determining the standards for the use of a term in a context – metalinguistic disagreement, in particular, is a shared speech situation in which public terms of conceptual engagement, the common practice of classifying, are manifested. Thus, such mechanisms must be resisted.

As said at the beginning of this chapter, one suggested means by which we can resist such mechanisms is by making interventions within conversation in which there is disagreement about concepts (Cantalamessa 2019; Sterken 2020). Metalinguistic discourse, or so the story goes, provides a space for contesting existing terms of conceptual engagement, allowing for the introduction, and subsequent propagation, of new classificatory dispositions. This is because metalinguistic discourse is a location at which disagreement about concepts can arise. And the possibility of disagreement, as a rational exchange of reasons, means that justification for the propagation of a new classificatory practice can be offered, and ultimately accepted in dominant contexts. This constitutes a particular individualistic conceptual implementation strategy.

In the next chapter, I will show that real-world power structures affect the distribution of control in the joint activity of pairing concepts with words in a way that tends to advantage those who endorse

dominant patterns of classification. That is, metalinguistic disagreements are often *non-ideal.* And, importantly, such disagreements tend to take place under conditions of injustice. Because of this, making interventions within metalinguistic disagreement, when one is at a disadvantage, means that convincing another to take up an ameliorative proposal, or else to get another to downgrade their confidence in extant dominant concepts, is highly unlikely (and sometimes costly). Social forces, such as norms and power dynamics, get in the way of successful individual conceptual advocacy.

# Chapter Five

## Non-Ideal Metalinguistic Disagreement

*When language is used to attribute properties to language or otherwise theorize about it, a device is needed that – to borrow Donald Davidson's apt phrase – turns language on itself.*

Herman Cappelen and Ernie Lepore 2007, p. 1

### 5. Introduction

The previous chapter explored particular mechanisms of conceptual reproduction that occur within metalinguistic discourse: *Establishment* and *Negotiation.* Both mechanisms are features of our everyday lives, and as such are regular and common means by which public terms of conceptual engagement are manifested. What's important to highlight is that such mechanisms often involve the reproduction of *ideological* concepts; those concepts that contribute to upholding or sustaining systems of unjust hierarchy. Because of this, it stands to reason that we should find ways to resist, disrupt, or else undermine such mechanisms in order to prevent ideological concepts from being reproduced and playing a role in structuring our lives together.

As mentioned at the end of the previous chapter, one kind of suggestion, a particular individualistic implementation strategy, that has been put forward as a means of resisting, disrupting, or undermining mechanisms of conceptual reproduction that occur in metalinguistic discourse is to make certain interventions *within* metalinguistic discourse (e.g. Cantalamessa 2019; Sterken 2020). The thought seems to be that because metalinguistic discourse provides a space in which disagreement about concepts can arise (i.e., metalinguistic disagreement), then the possibility of disagreement, as a rational exchange of reasons, means that a new conceptual practice, or way of classifying aspects of the world, can be offered, and ultimately accepted in dominant contexts. Or, in other words, the possibility of disagreeing over the concept that should be expressed by a word in a context offers a site at which dominant terms of conceptual engagement can be resisted: one can introduce an alternative concept, developed from within non-dominant situatedness, and pit it against a dominant one to display just how superior it is.

In this chapter, I will show that there are unrecognised social obstacles to implementation strategies of this kind, which make such strategies unlikely to be successful, and subsequently stifles conceptual innovation. Specifically, I will examine the illegitimate power involved in the accommodation of conceptual attitudes within metalinguistic discourse; specifically, cases of *Negotiation* (i.e., metalinguistic disagreement). Metalinguistic disagreement, I contend, can be more or less ideal. In non-ideal disagreements one speaker has less power in this conceptual tug-of-war. And in such cases, conceptual accommodation occurs as a result of an unequal distribution of power in the joint activity of pairing concepts with words that tends to favour or advantage dominant patterns of classification. Because metalinguistic disagreement tends to favour or advantage dominant patterns of classification, making successful interventions within metalinguistic disagreement, as someone who endorses justice-promoting concepts, especially concepts that have been developed

from within marginalized situatedness, is doubtful. More often than not will one fail to convince another to adopt ameliorated conceptual practice, or downgrade their confidence in an existing word-concept pair.

To put it differently, a problem for conceptual engineering is that justice-promoting concepts must be reproduced if they are to gain community-wide uptake. However, there are drags in metalinguistic disagreement that will tend to favour the reproduction of existing terms of conceptual engagement. This casts a large shadow of doubt over the friendliness of metalinguistic discourse as a site at which a non-dominant conceptual practice can take root in dominant contexts.[141]

One final note before I continue. Non-ideal metalinguistic disagreement has been critically overlooked in the philosophical literature. The significance of this must be stressed. Metalinguistic disagreements often, if not *always*, take place within real world power structures. Consider an example.

Suppose that Fil and Fig are both chefs, deciding on the items that should be on a menu. Fil is the sous-chef, Fig is the head-chef. Fil believes that animals aren't food, and Fig thinks otherwise:

Fil: Cows aren't food. They're persons, after all.
Fig: Of course cows are food. Look around you; they're on menus everywhere.

At this point, all seems well in this back-and-forth: Fil and Fig appear to be engaging in a paradigm metalinguistic negotiation. However, imagine how the exchange might unfold.

Fig: I'm sick of your animal advocacy. In this restaurant, cows are food!
Fil: Yes, chef.

We can see that there is a clear difference in what Fil and Fig are able to achieve in this disagreement. Namely, Fig has more influence in the joint activity of pairing concepts with words *in virtue of* her social position. Given that we are embedded in systems of hierarchy, non-ideal metalinguistic disagreements are common; plausibly more common than ideal exchanges. Thus, failure to account for non-ideal metalinguistic disagreement is a failure to explain how metalinguistic disagreements tend to take place out in the wild.

Not all non-ideal metalinguistic disagreements are ideological. Despite being hierarchical, some involve an unequal but non-problematic distribution of control over the joint activity of pairing concepts with words – perhaps this is true in the foregoing case. However, some non-ideal metalinguistic disagreements are deeply concerning. There are innumerable everyday cases in which, owing to identity prejudice, there is asymmetric control between parties to metalinguistic disagreement, limiting the metalinguistic moves certain speakers can make. This will be the subject matter of the following three chapters.

[141] This chapter raises concerns for the idea that metalinguistic negotiation is a site at which a new conceptual practice might propagate and spread across a community. However, there is a sense in which it also problematizes 'speaker-meaning' conceptual engineering (Pinder forthcoming). After all, this version of conceptual engineering requires that speakers are given an opportunity to have their speaker-meanings heard, accepted, and subsequently used in conversation. And Pinder even suggests that this is 'easy': all that is required to maximise the chances of success is for a speaker to be careful when articulating their definitions in a local context (forthcoming, p. 17). In time, the speaker-meaning might even be taken up more widely within a community. However, I will argue, this is idealistic. It fails to respect the real-world power relations that affect one's ability to impact conversational outcomes: speaker-meanings are often not taken seriously, ignored, silenced, etc. owing to the operation of identity prejudice in a context.

## 5.1. Metalinguistic Disagreement

When I apply a context-sensitive expression to an object, I can do one of two things: I can rely on the context to make sense of the properties of the object, or I can rely on the properties of the object to make sense of the context. Plunkett and Sundell (2013) call the latter *metalinguistic usage.* And when interlocutors exchange conflicting metalinguistic usages, they are engaging in *metalinguistic disagreement.* We might ask: What reasons does one have to pursue disagreement of this kind? Of course, there are many. I will take time to explore three prominent reasons, none of which are essential to metalinguistic disagreement.

One reason to pursue metalinguistic disagreement is to *complete* or *resolve* the joint activity of pairing concepts with words. Call this *metalinguistic negotiation.*[142] This isn't strictly the definition provided by Plunkett and Sundell (2013, p. 3). For Plunkett and Sundell, metalinguistic negotiation *just is* metalinguistic disagreement as I have described it above. Nevertheless, a distinction is warranted. After all, sometimes speakers might not have the goal of pairing concepts with words. For instance, imagine two philosophers who simply want to assess the relative merits of alternative concepts that could be expressed by 'woman' without coming to a conclusion. Moreover, the definition of metalinguistic negotiation that I have offered seems to appear implicitly in Plunkett and Sundell's discussion of a fictional disagreement between Oscar and Callie over the concept that should be expressed by 'spicy' (see also chapter four, §4.7). They argue the following:

> Why would Oscar and Callie consider it worth their time to engage in such a disagreement, when they already agree on what the chili actually tastes like? ... [B]ecause how we use words matters. For Oscar and Callie, as for many of us, an agreement amongst all the cooks in the kitchen that the chili can be described as 'spicy' plays an important role in... decision-making about whether to add more spice (ibid. 2013, p. 15).

In this case, the reason for pursuing metalinguistic disagreement is not simply for the sake of exchanging metalinguistic usages. Rather, the goal is to complete the process of pairing concepts with words. Particular to this example, it seems that the speakers have the goal of completing metalinguistic disagreement in order coordinate on the use of an expression, which will achieve a practical end that both speakers care about.

In this chapter, I will be primarily concerned with metalinguistic negotiation as I have defined it. However, I am not going to discuss cases in which speakers aim to achieve consensus in order to realize a shared practical interest. Moreover, I will not be so interested in cases where the goal metalinguistic negotiation, for both speakers, is to coordinate on the use of an expression. Rather, I will focus on situations in which the need for resolving metalinguistic disagreement is forced or imposed onto an interlocutor; typically, to realize an end, practical or otherwise, that only one speaker cares about.

Other theorists have offered alterative reasons to pursue metalinguistic disagreement. For example, Alex Davies (2020) argues that not all metalinguistic disagreements are deliberate exchanges that aim at resolution. Often, such disagreements are used as a platform on which identity

[142] How can we tell whether a disagreement is metalinguistic? Amie Thomasson (2017) offers some diagnostic criteria: (i) interlocutors agree on the facts, and no further facts would resolve the disagreement; (ii) the dispute is insensitive to prior or actual usage of the term; and (iii) the disagreement survives even with the recognition that terms are being used differently.

is displayed. When a word is context-sensitive, or polysemous, there are several criteria that one could use in a context. Engaging in metalinguistic disagreement, where one advocates for one or more criteria, serves as a means of 'giving off' information that isn't merely about getting an audience to recognize an intention to use a word in a certain way. Thus, for Davies, identities can be displayed through disagreement insofar as information can be conveyed about the social category to which one belongs, or through the performance of a ritual that expresses shared identity (ibid. 2020, p. 11).

Besides Davies, however, some have argued for the significance of pursing metalinguistic disagreement for the purposes of implementing an ameliorative strategy, or to get a hearer to downgrade their confidence in an existing word-concept pair.

Rachel Sterken (2020) has argued that engaging in metalinguistic disagreement can prompt an audience to counterfactually reflect on extant word-concept pairs. In her advocacy for this position, Sterken responds to the following challenge: deploying an improved word-concept pair, a product of conceptual amelioration, will hamper effective communication since the use of such a pair is a *linguistic transgression* – it involves speaking outside the language of a community.[143] However, Sterken argues that this is a feature of ameliorative analysis, not a bug. We needn't worry about deploying ameliorated word-concept pairs since it serves as a *linguistic disruption*, unsettling common ground so as to cause an audience to reflect on their linguistic decisions. The hope is that this will constitute a *transformative communicative disruption*. The hearer reflects on the ameliorated word-concept pair, recognizes it as a viable alternative, and sees it as an improvement.

Elizabeth Amber Cantalamessa (2019), in her important work on conceptual activism and disability studies, argues for something similar. However, Cantalamessa does not think that all disruptions must aim at causing another to adopt an ameliorated word-concept pair. One can simply encourage an audience to reassess their confidence in existing ones. The goal of which is to motivate speakers to give up their dominant and oppressive concepts, which will go some way towards overcoming certain forms of injustice.

I take such practical, social, and political reasons to be appealing. At the same time, however, I recognize that these reasons are not wholly sensitive to the realities in which metalinguistic disagreements tend to take place. Each appears to presuppose something about the participants to metalinguistic disagreement, and the social environment in which they are embedded. Specifically, the presupposition is that metalinguistic disagreement takes place under *ideal* or *near ideal* conditions:

> **Ideal Metalinguistic Disagreement:** A metalinguistic disagreement that takes place between *S* and *T*, such that *S* and *T* (1) are *metalinguistic peers* with (2) a high degree of *local metalinguistic agency*.

This is a mouthful. I will take time explicating the relevant concepts needed to understand this definition: *metalinguistic move, metalinguistic agency,* and *metalinguistic peerhood.* Further, I will show that not all metalinguistic disagreements are ideal. Many, if not most, are *non-ideal*:

> **Non-Ideal Metalinguistic Disagreement:** A metalinguistic disagreement that takes place between *S* and *T*, such that *either S* and *T* (1) are not metalinguistic peers, *or* (2) at least one does not have a high degree of local metalinguistic agency.

[143] Jennifer Saul (2006) has raised this concern for Haslanger's ameliorative definition of *woman.*

A note before I continue. I want to make it clear that I am not suggesting that each of the foregoing reasons to engage in metalinguistic disagreement cannot be present in *non-ideal* metalinguistic disagreement. My point is simply that there is a critical omission of discussion from theorists working on metalinguistic disagreement about the ways in which the kinematics of such disagreements are often influenced by the (illegitimate) power operating within a context. Elucidating this is my primary task.

### 5.2. M-Moves

An action counts as a move in an activity when it contributes to that activity (McGowan 2019, 86). The activity under discussion is metalinguistic disagreement. Thus, a *metalinguistic* move (m-move) is a conversational move that counts as a contribution to metalinguistic disagreement. Such moves involve two features: *type* and *impression*. An m-move type is the kind of speech act it is; the illocutionary force of a metalinguistic speech act. The impression of an m-move is the degree to which the speech act contributes to metalinguistic disagreement, its perlocutionary effect. Let's first explore the former.

Metalinguistic speech acts, or m-move types, are the kinds of speech acts one can perform in a metalinguistic disagreement. To date, no one has offered an exhaustive taxonomy of this category of speech acts. However, there are some examples. Nat Hansen has brought to light a range of m-move types that fall under the banner *metalinguistic proposals*: A speech act, which is a sub-category of advisories, where a speaker intends for an audience to come to have a reason to use an expression in a certain way (Hansen 2019, p. 1). There could be others. Metalinguistic *assertions* could be understood as speech that involves asserting that a particular concept is paired with a word as if it were true (e.g., 'Non-human animals are not food'). Metalinguistic *questions* could be understood as speech that involves asking which concept is paired with a word. And, it is plausible that metalinguistic speech acts are often *indirect*. After all, some speech acts can have two illocutionary forces.[144] When someone asserts, 'Non-human animals aren't food,' this could be understood as an object-level assertion *and* a meta-level proposal, or request, or demand, etc. for the use of a word in a context.

Proposals, it seems, do not require holding a certain kind of *authority* to be felicitous. Plausibly, however, the felicitous performance of other m-moves depends on contingent environmental facts, such as having authority over a hearer.[145] According to Nat Hansen, *metalinguistic directives* are one such category (2019, p. 3). Within the category of directives, there are: *metalinguistic requirements,* speech that *requires* a hearer to use a word-concept pair; *metalinguistic prohibitives,* speech that *forbids* a hearer from using a certain word-concept pair; and *metalinguistic permissives,* speech that *allows* a hearer to use a word-concept pair. (ibid. p. 3). Despite this, it is unclear that we need to make a strong commitment to authority being a constitutive condition for such speech acts. Perhaps someone without authority could felicitously perform a metalinguistic directive. However, those without authority will typically not be able to have a desired effect on conversation. Authority might have more to do conversational outcomes rather than the kind of speech act one can perform. This takes us to the *impression* of an m-move.

---

[144] For a detailed analysis of indirect speech acts, see Jörg Meibauer (2019).

[145] Making them a sub-category of *authoritative illocutions* (Langton 1993, p. 305).

The impression of an m-move is the degree to which the move contributes to the activity of metalinguistic disagreement. This is roughly the perlocutionary effect on conversation about word-concept pairs. And the degree to which an m-move contributes to metalinguistic disagreement depends on the type of disagreement in question. For metalinguistic negotiation, the impression of an m-move is its contribution to resolving a dispute; for identity-display, the impression of an m-move is its contribution to having one's identity recognized; for linguistic disruption, the impression of an m-move is its contribution to convincing another to (minimally) downgrade their confidence in an extant word-concept pair.

My primary interest is in the relationship between authority and impression. I won't take a stance on whether certain m-moves require authority to be felicitously performed. Rather, I accept that such moves might require authority, but nevertheless my concern is with the impact one can make on conversational outcomes.

To see the distinction between type and impression more clearly, think back to the exchange between Fil and Fig. Let's imagine Fil believes that, in the broader linguistic community, 'food' should be paired with a concept that does not classify non-human animals as food. Fil also takes the moralized, ameliorated 'food'-concept pair to make sense of his social identity (i.e., vegan). This informs the type of the m-moves that Fil wants to make (i.e., metalinguistic proposals) and their intended impression. He wants Fig to know that he is vegan; he wants Fig to downgrade her confidence in the extant, less moralized 'food'-concept pair; and he has beliefs about the sort of m-moves that will resolve the disagreement in his favor.

Recall how the exchange unfolded. Fed up with Fil's animal advocacy, Fig says, 'In my restaurant, cows are food!' We can see that this type of utterance is a metalinguistic requirement: Fig has performed a speech act requiring Fil to use a specific 'food'-concept pair. And, the impression is strong. The possibility of losing his job means that Fil must now use a specific 'food'-concept pair, at least in the restaurant. Fig has made a profound impact on conversational outcomes. She has exercised power to affect the character and direction of conversation in a way that is less available – if available at all – to Fil.

With the concept of m-move on the table, let's introduce *metalinguistic agency* (m-agency) and *metalinguistic peerhood* (m-peerhood).

## 5.3. M-Agency

M-agency is a scalar concept. It refers to the m-moves one can perform and the overall impression of such moves. There are a number of ways that m-agency can be affected. One can have *more* m-agency in virtue of being able to perform more moves or being able to perform moves that leave a stronger impression. One can have *less* m-agency in virtue of being able to perform fewer moves or being able to perform moves that leave a weaker impression.[146] We might ask: *More* or *less* relative to what?

One answer is that the degree of m-agency one possesses is determined by the type and impression of m-moves one can perform *relative to an interlocutor.* Call this *local m-agency.* Consider the foregoing exchange. Fig can perform m-moves with a stronger impression, relative to Fil. She's the boss, he is the subordinate.

Another answer is that the degree of m-agency one possesses is determined by the type and impression of m-moves one can make *relative to a range of contexts.* Here, m-agency concerns what

[146] How should we understand the m-agency of someone who has the ability to perform fewer m-moves, but with a stronger impression? I leave this question open.

one can do *throughout* social space. It refers to the m-moves a speaker can perform, and the impression of such moves, across a set of possible and actual conversations, and in virtue of the variable (e.g., gender, race, etc.) that one is enabled or constrained. Call this *global m-agency*.

Local and global m-agency are explanatorily interdependent. We cannot make sense of the typical range and impression of m-moves that a speaker can make without having a sense of what the speaker can do relative to a specific interlocutor; and we cannot make sense of a speaker's global m-agency by examining a specific conversation alone. Moreover, one's local m-agency may not always reflect what one can typically achieve in disagreement. There are times at which one might strong-arm their way into a dominant metalinguistic position, but this would be unusual across the board.

I am interested in what one can do in a specific conversation. For this reason, my interest lies in local m-agency. However, I am also concerned with what it is in virtue of that one's local m-agency is enabled or constrained. An examination of this requires understanding broader social practices that influence conversational kinematics. Specifically, I am interested in the role that *identity prejudice* plays in shaping the character and direction of metalinguistic disagreement. Thus, I am also interested in global m-agency. If identity prejudice plays a role in metalinguistic disagreement, this should affect a speaker across a variety of conversational contexts.

### 5.4. M-Peerhood

M-peerhood is specific to a conversation. It is a feature of metalinguistic disagreement when interlocutors have available to them the ability to perform the *same* range of m-moves (or roughly the same), with a similar degree of impression. In other words, participants to metalinguistic disagreement are *m-peers* when there is no m-move[147] that one participant can perform that the other cannot, and neither participant can leave a stronger impression with the same m-move. Further, m-peerhood does not assume a degree of m-agency possessed by interlocutors. Participants to metalinguistic disagreement can have low levels of local m-agency yet still be m-peers.

### 5.5. Ideal and Non-Ideal

Earlier, I suggested that the reasons offered in existing literature as to why one might pursue metalinguistic disagreement are not wholly sensitive to the realities in which metalinguistic disagreements tend to take place. The kinds of metalinguistic disagreements commonly explored are those that *ideal* or *near ideal*. We are now in a position to properly understand what this means:

> **Ideal Metalinguistic Disagreement:** A metalinguistic disagreement that takes place between *S* and *T*, such that *S* and *T* (1) are m-peers with (2) a high degree of local m-agency.

Put differently: Metalinguistic disagreement is ideal when participants to conversation can make the same extensive range of m-moves with the ability to leave (roughly) the same impression on conversation with the performance of such moves. Why think that theorists presuppose this?

Consider metalinguistic negotiation. Take an example that is representative of the paradigm cases that interest Plunkett and Sundell (2013). Thinking back to the previous chapter, suppose two friends are arguing over the status of a jacket:

[147] Or no importantly different m-move.

George: That jacket isn't smart casual. Black is too formal.
Kelly: But it doesn't even have shoulder pads!

We can see that the friends are not arguing over facts about the jacket – they agree on what it looks like. Instead, the dispute centers around whether the jacket should count as *smart casual*. George believes that because the jacket is black, it is too formal; Kelly believes that this isn't a problem because it lacks shoulder pads. And, in this case, George and Kelly are aiming at consensus. Kelly will wear the jacket depending on what is decided.

What makes this exchange ideal? First, George and Kelly are m-peers. There is no m-move, or no importantly different m-move, that is only available to either George or Kelly; and both can perform such moves with a similar degree of impression on conversation. [148] Second, George and Kelly have a high degree of local m-agency. Both are able to perform a wide range of m-moves, with a strong impression, because the conversation is between friends (and *vice versa*).[149]

Not all metalinguistic disagreements are this pleasant. Each condition can fail. Sometimes speakers are not m-peers; and sometimes a speaker does not enjoy high local m-agency. Further, such failures are often interrelated. The fact that speakers are not m-peers can mean that the local m-agency of one party to conversation is severely constrained. To see this, recall the exchange between chefs:

Fil: Cows aren't food. They're persons, after all.
Fig: Of course they're food! Look around you; they're on menus everywhere.

At this point, the interaction appears to be ideal. Fig and Fil seem like m-peers and both seem to enjoy high-levels of local m-agency. However, remember how the disagreement unfolded:

Fig: I'm sick of your animal advocacy. In this restaurant, cows are food!
Fil: Yes, chef.

This reveals something important: Fig and Fil are not m-peers. That is, Fig is able to have a more profound impact on the outcome of disagreement over which 'food'-concept pair will be operative in the context. In other words, Fig can perform a metalinguistic directive that will greatly affect the character and direction of the disagreement. One consequence of Fig's speech is that it delimits the local m-agency of Fil.[150] He is unable to freely express attitudes about his preferred 'food'-concept pair without the risk of punishment. In sum, the joint activity of pairing concepts with words is heavily skewed in Fig's favor. [151]

This example shows that the conditions for ideal metalinguistic disagreement can fail. The speakers are not m-peers; and at least one speaker does not enjoy high local m-agency. From this, we can define *non-ideal* metalinguistic disagreement as follows:

[148] Some friends may not be m-peers, but a *typical* friendship involves the ability to freely express oneself.

[149] There could be m-moves that are inappropriate for George and Kelly to perform owing to their friendship, such as metalinguistic directives.

[150] Though speakers may not be m-peers, both could nevertheless enjoy high-levels of local m-agency.

[151] Further, Fil cannot (easily) display his identity nor attempt a linguistic disruption without risking his job.

**Non-Ideal Metalinguistic Disagreement:** A metalinguistic disagreement that takes place between *S* and *T*, such that *either S* and *T* (1) are not m-peers, *or* (2) at least one does not have high local m-agency.[152]

Metalinguistic disagreements are a feature of everyday life. And given that power structures dictate almost all areas of our social existence, non-ideal exchanges are plausibly more common than ideal ones. The case above might not strike one as particularly important. And it must be made clear that not all non-ideal metalinguistic negotiations are unjust. Some involve a fair but unequal distribution of control over the joint activity of pairing concepts with words. Perhaps this is even true in the foregoing case. However, there are many instances of non-ideal metalinguistic disagreement to which we should be especially attentive. Consider the following:

Woman: You brushed up against my butt, that's sexual harassment!
Man: Don't be so sensitive, that's way too trivial to count.

This disagreement concerns the concept that should be expressed by 'sexual harassment.' And, we should be deeply worried about how exchanges of this kind unfold. Settling on a 'sexual harassment'-concept pair has significant practical and psychological consequences. If the man is able to have his preferred word-concept pair accepted by the woman, certain things follow. The woman must update her belief that sexual harassment didn't occur. She might feel the need to apologise for accusing the man of sexual harassment. She might second-guess herself when touched again. She might feel over-sensitive, hysterical, and paranoid. She might force herself to work in an unsafe environment. Etc.

The task of the rest of this chapter is to examine cases of non-ideal metalinguistic disagreement that *are* unjust. In particular, I will explore the role of power and identity in determining who has greater local m-agency in a context. I will argue that the joint activity of pairing concepts with words can be corrupted by the operation of *identity prejudice*.

What is identity prejudice? Ordinary understanding tells us that 'prejudice' is an aspect of individual psychology. *People* are prejudiced when they have an unfounded attitude. This over-psychologizes what is a more useful concept. Prejudice, as I will understand it, is a property of social environments. It is a feature of social structure. And, it involves how one's actions and thoughts are enabled or constrained in virtue of belonging to a particular social category.[153] *Identity prejudice*, then, is a structural phenomenon in which one's actions and thoughts are enabled or constrained owing to the identity category to which one belongs. This includes gender, race, ability, class, etc. For example, think of a predominantly Black neighbourhood in which there is a non-accidental absence of polling booths for a presidential election. This is a racist social environment that is prejudiced against a particular identity category (i.e., Black Americans). Thus, when I say that I am interested in exploring the role that identity prejudice plays in metalinguistic negotiation, I am not (just) interested in the attitudes of bad eggs. I am concerned with properties of the external world.[154]

Some notes before I continue. It is plausible that an illegitimate difference in m-agency can be unjust, unfair, harmful, wrongful, etc. without involving identity prejudice. Given this, I want to

[152] This definition is inclusively disjunctive. And, the disjuncts are scalar, which means that a metalinguistic disagreement can be *more* or *less* non-ideal.

[153] Kate Manne (2018) says something similar in her account of misogyny.

[154] This comes close to Fricker's notion of *structural identity prejudice* (2007, p. 155).

explicitly state that my interest is in forms of identity-based oppression. I leave it open as to how one might spell out non-identity related wrongs of particular non-ideal metalinguistic disagreements.

Going forward, I will focus on non-ideal metalinguistic *negotiation.* In particular, I will examine situations in which the need to complete or resolve metalinguistic disagreement is forced or imposed onto an interlocutor.

## 5.6. Metalinguistic Power

Metalinguistic negotiation, as I have understood it, involves speakers aiming to complete the joint activity of pairing concepts with words. Moreover, metalinguistic negotiations can be (more or less) ideal or non-ideal. I will focus on *non-ideal metalinguistic negotiation*. Specifically, I am interested in what makes metalinguistic negotiation non-ideal and whether this can be unjust.

In the joint activity of pairing concepts with words, one speaker can have more or less control over this process. Often, this control is not equally distributed across speakers. One can have greater influence over the concept that a word will express in a context. This is the hallmark of non-ideal metalinguistic negotiation. However, this difference in control can be innocuous. Not all non-ideal metalinguistic negotiations with unequal control are unfair. After all, we should expect a speaker to defer to another on the grounds of epistemic expertise – it seems that an intern ought to accept the preferred word-concept pair of an attending doctor, at least under normal conditions.

Despite this, unequal control in metalinguistic negotiation is often a function of being situated in unjust systems of dominance and subordination, such as hierarchical relations between members of certain identity groups (e.g., White/Black). When one has greater control in the joint activity of pairing concepts with words *in virtue of* belonging to an identity group that occupies a position of unjust dominance over the identity group of an interlocutor, then this control is illegitimate. Call this *metalinguistic power.* [155] Metalinguistic power is the possession of unjust control in metalinguistic negotiation that is a function of interacting parties from social categories implicated in oppressive social relations. To explore this further, let's return to the foregoing example:

Woman: You brushed up against my butt, that's sexual harassment!
Man: Don't be so sensitive, that's way too trivial to count.

To reiterate, this disagreement concerns the concept that should be expressed by 'sexual harassment.' The man thinks that his intentional action isn't sexual harassment; the woman thinks that it is. Suppose the man gets his way in this exchange. I argue that we can understand this as occurring in virtue of the man having metalinguistic power over the woman; a power derived from the operation of identity prejudice in the context. But to see this more clearly, we need to know about the *social environment* in which the exchange takes place.[156]

## 5.7. Social Environments

[155] I am using 'power' as *power-over* (Weber 1978).

[156] One might worry about my reliance on fictional cases. However, there are many cases out in the wild that my analysis makes sense of. Teresa Marques (2020) argues that the expression of 'free election' has a positive connotation. However, certain politicians often use the term not as a means of protecting the rights of citizens, but instead to manipulate people into thinking that particular 'elections' involve fair or just processes. This difference in usage constitutes a type of (unjust) power-affected metalinguistic disagreement.

A social environment is a network of interrelated and regular patterns of coordinated social behaviour (Podosky forthcoming). Such patterns depend on the culturally available information that agents draw on, such as characterizations, tropes, narratives, social meanings, schemas, roles, etc. This information frames expectation, and serves to stabilize behaviour by providing rules that govern social interaction. Agents rely on such information to render intelligible experience, which then facilitates intentional engagement and coordination with others and surroundings. Moreover, *social environments can be unjust*. This occurs when the extant patterns of coordinated behaviour unfairly privilege some and subordinate others. Consider an example.

During a visit to the hospital, a woman finds that her testimony is not being taken seriously, leading to a misdiagnosis. This owes to the salience of a particular schema 'woman' that triggers prejudicial stereotypes (or generic characterizations) associated with women, such as being 'too sensitive.' That is, the doctor engages in a pernicious pattern of deflating his assessment of the woman's credibility by drawing on defective culturally available information that he uses to make sense of the situation, and to guide his decision-making practices.

With respect to the sexual harassment case above, we can tell a similar identity prejudicial story. Let's imagine that when the woman accuses the man of sexual harassment, she does so in the context of a male dominated workplace, situated in a patriarchal social environment. Here, gender schemas are salient: A *woman* has accused a *man* of sexual harassment in a workplace where men are perceived as better qualified, objective, and more legitimate. When the man retorts, 'Don't be so sensitive,' it should not come as a surprise that the woman downgrades her self-trust, and subsequently comes to doubt the accuracy or aptness of her 'sexual harassment'-concept pair. The man has gaslighted the woman through controlling the outcome of metalinguistic negotiation. I will explore this in more detail in chapter seven.

What we can see is that the culturally available information of a patriarchal social environment corrupts epistemic judgments (Fricker 2007). Women are often, and incorrectly, perceived as over-sensitive, paranoid, hysterical, unreasonable, subjective, etc., which subsequently subjects them to persistent and pernicious challenges to their epistemic reliability (e.g., 'Did you lead him on?' etc.). Because of this, women tend not to be believed when they testify to sexual harassment (or sexual assault). Thus, it is 'fitting' for women to downgrade their self-trust in the face of such challenges. It is what is expected and demanded of them, especially in a workplace where men are perceived as more reasonable.

Going back to our example, we might begin to see how the man comes to have metalinguistic power over the woman. Given the social environment, the man unfairly occupies a position of epistemic dominance. And this owes to the *identity* of each speaker and the prevailing social stereotypes that affect coordination of cognition, affect, and practice. He is a man, and men are trusted sources of information. She is a woman, and women are irrational and unreliable. In sum: the power to affect metalinguistic disagreement is distributed unequally in society. Importantly, this unequal distribution is often along identity categorial lines.

## 5.8. Metalinguistic Power, M-Peerhood, and M-Agency

How does metalinguistic power relate to the discussion above? *The presence of metalinguistic power in the joint activity of pairing concepts with words entails non-ideal conditions.*[157] When one has metalinguistic power, one has more control in the joint activity of pairing concepts with words. When one has more control, one is not an m-peer with their interlocutor. And when one is able to do more with their words, the local m-agency of the other is constrained.

In the foregoing case, metalinguistic power has more to do with the *degree of impression* one can have on conversation with the performance of an m-move, rather than the types of m-moves one can perform. The impression of the man's m-moves, in resolving metalinguistic negotiation, is far stronger than the woman's. He is more able to profoundly impact conversational outcomes. Thus, the woman's local m-agency is constrained. She is unable to perform m-moves with the same impression. And this constraint owes to her being a woman situated in a patriarchal social environment. Her words carry less weight *in virtue of belonging to a particular identity group.*[158]

## 5.9. Conclusion

When metalinguistic power is present in conversation, the dominant speaker can exercise control in a way that allows them to coercively impose their preferred concept on an interlocutor. Subsequently, the concept is reproduced in that context. And if the concept is a part of the collectively shared set of coordinated concepts within a linguistic community, then this further counts as contributing to global conceptual reproduction. This helps sustain existing terms of conceptual engagement, which undergirds and maintains our behavioural dispositions.

If the concepts that are endorsed by dominant speakers are ideological, those concepts that undergird oppressive social practices, and therefore contribute to sustaining unjust hierarchy, then the presence of metalinguistic power is a real problem for theorists who believe that metalinguistic negotiation is a location at which implementing an ameliorative proposal can be successful; such as Sterken (2020) and Cantalamessa (2019). To reiterate, Sterken believes that the use of an ameliorated word-concept pair will not be an issue for communicative purposes since it can cause a transformative communicative disruption – that is, the use of ameliorated word-concept pair will make a hearer counterfactually reflect on their language, and prompt them to take up the new concept. Cantalamessa argues for something similar. However, for Cantalamessa, the goal needn't be to get another to take up a word-concept pair, but rather to simply cause another to downgrade their confidence in existing pairs.

Why is non-ideal metalinguistic negotiation a problem for theorists such as Sterken and Cantalamessa? The resolution of metalinguistic negotiation will tend to favour dominant speakers who endorse existing terms of conceptual engagement. This makes the propagation of an ameliorative proposal, within metalinguistic discourse, highly unlikely. Further, it will also reduce the

[157] Note that non-ideal conditions do not entail the presence of metalinguistic power. This is because a speaker can have less control in the joint activity of pairing contents with words without being subject to identity prejudice.

[158] There are other ways that metalinguistic power can shape the character and direction of metalinguistic disagreement beyond what I have suggested. For instance, a powerful speaker can *steamroll* a marginalized hearer insofar as the powerful speaker can force the completion of metalinguistic negotiation onto a hearer; or a powerful speaker might simply be *apathetic* insofar as a marginalized speaker wants to complete metalinguistic negotiation but the powerful speaker is unwilling; or a powerful speaker might simply *pressure* a marginalized hearer to reassess their confidence in an existing word-concept pair, without aiming to resolve metalinguistic negotiation; etc.

chances of convincing another to downgrade their confidence in existing word-concept pairs. This is especially true for concepts that have been developed from within marginalized situatedness, and that directly resist or challenge ideological terms of conceptual engagement. That is, marginalized concepts, especially when endorsed by marginalized speakers, stand very little chance of overcoming the conservative dynamics of metalinguistic disagreements in which (unjust) power is operative. The joint activity of pairing concepts with words in a context is geared towards advantaging those who advocate for dominant patterns of classification.

Where does this leave us? I believe that we should look beyond making interventions within metalinguistic discourse as a means of implementing conceptual engineering projects. Of course, such interventions should be tried. If someone is using a pernicious ideological concept in conversation, then perhaps the best defensive strategy is to challenge the concept, and shift the exchange into metalinguistic disagreement. However, defence is one thing, and resistance is another. The latter, I take it, is concerned with overturning the status quo – not simply protecting oneself, or others, against it. And in order to overturn the conceptual status quo, or the dominant terms of conceptual engagement, I believe that we must look beyond individualistic implementation strategies that focus on conceptual advocacy. The social infrastructure that protects that conceptual status quo is often far too hard to compete with as isolated individuals. Exactly which strategies we need goes beyond the scope of the thesis. I leave such concerns open for further research.

What I want to focus on in the final three chapters of this thesis is the idea that when identity prejudice plays a role in constraining what one can achieve in metalinguistic discourse, that this constitutes distinctive wrongdoings. So, not only are we competing with structures that work to keep things as they conceptually are, we are also battling structures of oppression. In the next chapter, I will explore a phenomenon that I call *metalinguistic injustice.* In particular, I will examine possible ways of understanding its wrong-making features. I will argue that metalinguistic injustice is either an *epistemic injustice,* a *linguistic injustice*, or both.

# Chapter Six

## Metalinguistic Injustice

*The dominated live in a world structured by others for their purposes – purposes that at the very least are not our own and that are in various degrees inimical to our development and even existence.*

Nancy Hartstock 1998, p. 241

### 6. Introduction

In the previous chapter, I argued that often non-ideal metalinguistic disagreements have illegitimate power running through them. This means that participants to such disagreements do not have equal control in the joint activity of pairing concepts with words – and, importantly, such disagreements tend to favour dominant patterns of classification. In this chapter and the next, I will explore how this mechanism of conceptual reproduction, or non-ideal metalinguistic disagreement, occurs under conditions of injustice. That is, I will introduce and examine a distinctive wrong that occurs in cases of metalinguistic disagreement in which there is an *unfair* or *unjust* distribution of metalinguistic control owing to the operation of identity prejudice. Such prejudices have the effect of delimiting the possible moves one can make in the joint activity of pairing concepts with words. I call this *metalinguistic injustice.* Precisely what constitutes metalinguistic injustice is spelled out. I argue that it can either take the form of an epistemic injustice, a linguistic injustice, or both simultaneously. The wrong (either) involves restricting someone from participating in the processes that determine the epistemic or linguistic resources of a conversation, and/or undermining someone's ability to affect metalinguistic outcomes. However it is construed, metalinguistic injustice makes implementing an ameliorative strategy within metalinguistic discourse highly unlikely.

I have already discussed how it is that metalinguistic power relates to both conceptual reproduction and the implementation challenge of conceptual engineering in the previous chapter. Given this, I won't have much to say about this relationship in what's to come. Metalinguistic injustice is simply what occurs when metalinguistic power is present. What I will discuss, not as a matter of implementation, but rather as a matter of defending oneself from pernicious ways of thinking, are corrective measures against metalinguistic injustice. I will not be able to offer a comprehensive account, but I will gesture at some possibilities. Such possibilities are limited to those who are motivated to avoid perpetuating moral wrongs. I understand that this is limited, but nevertheless it will go some way towards protecting marginalized folk from the threat of accidental reproduction of ideological concepts.

Before I continue, I want to reiterate what it means for identity prejudice to be operative in conversation. As said in the previous chapter, identity prejudice is not an aspect of individual psychology. It is not the attitude of bad eggs. Rather, as I have understood it, prejudice is a property of social environments – a network of interrelated and regular patterns of coordinated social behaviour. What does this mean? Prejudice concerns how one's actions and thoughts are constrained given the environment in which they are situated. *Identity* prejudice, then, is about being constrained

in virtue of belonging to a particular identity category (i.e., woman, man). Specifically, I am interested in the role of identity prejudice in metalinguistic disagreement. That is, how one is constrained (or enabled) in their ability to contribute to disagreement with respect to which concept will be expressed by a shared word, or set of words, in a context.

### 6.1. The Pooling of Cognitive Resources

As we all probably know by now, *epistemic injustice* occurs when one is undermined in their capacity as an epistemic subject (Fricker 2007). According to Miranda Fricker, forms of epistemic injustice have something in common: 'prejudicial exclusion from participation in the spread of knowledge' (2007, p. 162). For *testimonial injustice,* the primary harm involves 'exclusion from the pooling of knowledge owing to identity prejudice on the part of the hearer'; and the primary harm of *hermeneutical injustice* involves 'exclusion from the pooling of knowledge owing to structural identity prejudice in the collective hermeneutical resource' (ibid. p. 162). As a general characterization, epistemic injustice can involve either identity prejudice in the hearer *or* identity prejudice in the shared cognitive resources, such as shared classificatory practices and overlapping characterizations.

I will explore whether metalinguistic injustice can be construed as involving either of these forms of epistemic exclusion. Further, I will examine whether metalinguistic injustice constitutes a *linguistic injustice.* This is a critically under-theorized notion. Similar to epistemic injustice, linguistic injustice involves prejudicial forms of exclusion that involve identity prejudice in the hearer or identity prejudice in the shared linguistic resources (i.e., classificatory practices).

Overall, my aim will be to assess whether metalinguistic injustice constitutes (i) identity prejudice in the hearer or (ii) identity prejudice in the pooling of local resources. If metalinguistic injustice satisfies (i) or (ii), then it is either an epistemic injustice, a linguistic injustice, or both simultaneously. I will start by exploring whether metalinguistic injustice can be understood as an epistemic injustice; specifically, whether it can be understood as involving identity prejudice in the pooling of local cognitive resources.

To suffer from epistemic injustice is to be undermined in one's capacity as an epistemic subject (Fricker 2007; Gerken 2019). Note that this generalizes its original formulation. Previously, epistemic injustice was characterized as 'a wrong done to someone in their capacity as a *knower*' (Fricker 2007, p. 1). This change is important. The primary reason for this shift in terminology is to make explicit that our epistemic lives consist of understanding that fails to meet the strict conditions of knowledge. In other words, there are many ways to be an epistemic subject.

One way to be an epistemic subject is to be the kind of thing that interprets and represents the world on the basis of the available cognitive resources – such as coordinated concepts. Miranda Fricker expounds on this idea in her discussion on *hermeneutical injustice* (2007, p. 148). She argues that an epistemic subject renders intelligible experience with the cognitive resources at their disposal, and when one is denied access to such resources, under certain conditions, this constitutes a distinctively epistemic wrong.

More than this, according to Fricker, an epistemic subject is not just one who passively receives information, but also actively contributes to the set of cognitive resources shared between members of a thinking community. This idea is especially clear in the literature on pernicious ignorance (Pohlhaus Jr. 2012; Dotson 2012). Here, the concern is that dominantly situated epistemic subjects have far too much control over the cognitive resources that are available to members of a community of thinkers, and this is, partly, to do with the fact that there is systematic and wilful failure to recognize

the cognitive resources developed from within marginalized situatedness. Thus, we can say that an epistemic subject is, *inter alia,* the kind of thing that is *entitled* to contribute to the cognitive resources of a community of thinkers.

Cognitive resources do not just circulate within global communities. They are a feature of local contexts, such as conversation. If one way to be an epistemic subject is to be the kind of thing that is entitled to contribute to the cognitive resources of a global community, it is no stretch to say that another way to be an epistemic subject is *to be the kind of thing that is entitled to contribute to the cognitive resources of a local context.*[159] And, as we have explored, the joint activity of pairing concepts with words, metalinguistic negotiation, can be skewed in favour of one speaker. Thus, just as groups are constrained in their ability to contribute to the available cognitive resources in global settings, individuals can be constrained in their ability to contribute to the available cognitive resources in local contexts. Call this *metalinguistic deprivation*:

> **Metalinguistic deprivation** occurs when one is restricted, relative to their interlocutor, in their ability to contribute to the processes involved in determining the concept that will be expressed by a word, or set of words, in a context.

Metalinguistic deprivation is close to Fricker's notion of *hermeneutical marginalization* (2007, p. 152). This occurs when marginalized group members are prevented from contributing equally to the cognitive resources of a community, and therefore unable to render intelligible certain experiences, and cannot communicate such experience across social space. Thus, metalinguistic deprivation and hermeneutical marginalization both involve restricting people from participating in the processes that give rise to shared cognitive resources. There is a clear and important difference, however. Hermeneutical marginalization is the restriction of groups. Metalinguistic deprivation occurs between dyads. Another difference is that hermeneutical marginalization is a structural phenomenon that privileges certain groups, and disadvantages others. This isn't built into metalinguistic deprivation. Because of this, metalinguistic deprivation fails to constitute an epistemic *injustice.* Merely being prevented from contributing to the joint activity of pairing concepts with words can be innocent. Think of the case of the chefs, Fil and Fig, in the previous chapter. In this case, Fil is the sous-chef, Fig is the head-chef. They are discussing what items should be on the menu. Fil believes that cows should not be on the menu because they are not 'food'; Fig thinks otherwise. Arguably, it seems that Fil should have less control in determining which social concept of food is apt or appropriate for the context. Fil is only the sous-chef, after all. Thus, we can ask: What makes metalinguistic deprivation an epistemic injustice?

Consider the disagreement in the previous chapter between the man and woman over which concept should be expressed by 'sexual harassment.' In this case, the woman accommodates the man's conceptual attitude that a particular 'sexual harassment'-concept pair is apt or appropriate for the context. This is due to the man's metalinguistic power and the woman's metalinguistic deprivation. Importantly, her words carry less weight in virtue of *being a woman* and his words carry more weight in virtue of *being a man*. What we can see is that the operation of identity prejudice is responsible for the woman having less say, and the man having more say, in metalinguistic negotiation. That is, the woman and man are embedded in a context that is patriarchal, sexist, and

[159] Not *merely* contributing in a minimal sense but to a significant degree, where the level of significance is necessarily vague.

misogynistic, and these environmental factors shape the direction and character of metalinguistic disagreement in favour of the man.

With this in mind, how does metalinguistic deprivation constitute an epistemic injustice? In the first place, one has to be *entitled* to contribute to the cognitive resources of a local context. In the case that we've been considering, the woman is entitled (more so than the man) to contribute to the cognitive resources of the local context concerning which social concept should be expressed by 'sexual harassment.' Second, one has to be restricted in their ability to contribute to the processes that determine which concept will be expressed by a shared word in a context – they must suffer from metalinguistic deprivation. And third, they must suffer from metalinguistic deprivation *in virtue of* the operation of identity prejudice in the context (i.e., metalinguistic power). For example, the woman is unable to contribute equally to the processes that determine the relevant 'sexual harassment'-concept pair *because she is a woman.* When metalinguistic deprivation constitutes an epistemic injustice, call it *metalinguistic injustice (EI)*:

> **Metalinguistic Injustice (EI)** occurs when: (i) One is entitled to contribute to the cognitive resources of a local context, but (ii) one is restricted in their ability to participate in the joint activity of pairing concepts with words (i.e., metalinguistic deprivation), in virtue of (iii) the operation of metalinguistic power in the context.

### 6.2. The Pooling of Linguistic Resources

Apart from being an epistemic injustice, metalinguistic injustice may constitute a *linguistic injustice.* Following Ishani Maitra (2017), we can understand linguistic injustice in different ways. On one view 'a linguistic injustice may be an injustice committed against someone in their capacity as a linguistic agent, e.g., a speaker or hearer... And on yet another view a linguistic injustice may be unfair exclusion from a linguistic community... And there may be further options as well' (Maitra 2017, p. 288).

The wrong present in some forms of non-ideal metalinguistic negotiation could be captured by either suggestion. Metalinguistic negotiation could be wrongful because: (1) a speaker is undermined in their capacity as a linguistic subject in virtue of being unable to perform certain m-moves, or being unable to perform m-moves that leave a strong impression on conversation, or (2) a speaker is prevented from contributing to the pool of linguistic resources that would enable to them to accurately describe their experience. I will first examine (2).

(2) appears to be the kind of wrong that Fricker calls *hermeneutical injustice*, a species of epistemic injustice (2007, Ch. 5). However, Maitra suggests that there are grounds for thinking that the injustice here is distinctively *linguistic*, not epistemic. The reason for this is because, owing to hermeneutical marginalization, the 'agent suffering this wrong is prevented from accurately *describing* their own experience, and as a result, from *communicating* the nature of that experience to others' (2017, p. 289 my emphasis). For Maitra, describing and communicating are paradigm linguistic capacities.[160]

Importantly, we can connect these thoughts to our discussion above. Instead of thinking of metalinguistic injustice as concerning contribution to *cognitive resource*s, we can understand it as concerning contribution to *linguistic* resources. Drawing similarities to metalinguistic injustice (EI),

[160] We can say that hermeneutical injustice is *both* an epistemic injustice and a linguistic injustice.

we can say that metalinguistic deprivation constitutes a linguistic injustice as follows. One has to be *entitled* to contribute to the *linguistic* resources of a local context (e.g., the concept expressed by 'sexual harassment'); one has to be restricted in their ability to contribute to the processes that determine the concept that will be expressed by a shared word in a context (i.e., metalinguistic deprivation); and one must be subject to metalinguistic deprivation *in virtue of* the operation of identity prejudice in that context. When metalinguistic deprivation constitutes a linguistic injustice, in the sense we've been considering, call it *metalinguistic injustice (LI)*:

> **Metalinguistic Injustice (LI)** occurs when: (i) One is entitled to contribute to the linguistic resources of a local context, but (ii) one is restricted in their ability to participate in the joint activity of pairing concepts with words, in virtue of (iii) the operation of metalinguistic power in the context.

### 6.3. Prejudice in the Hearer: Testimonial Injustice

I want to move away from thinking about the wrong of metalinguistic injustice as prejudicial exclusion from the pooling of resources, be it epistemic or linguistic. I will now focus on whether metalinguistic injustice involves identity prejudice in the *hearer*.

For Fricker, when one's basic epistemic capacities are undermined, this constitutes 'epistemic objectification' (2007, p. 132 – 133). This draws on a distinction between *informants* on the one hand, and *mere sources of information* on the other:

> Broadly speaking, informants are epistemic agents who convey information, whereas sources of information are states of affairs from which the inquirer may be in a position to glean information. Thus, while objects can only be sources of information, people can be either informants... or sources of information... (ibid. p. 132).

Fricker argues that treating someone as a *mere* source of information constitutes epistemic objectification. It is a denial of epistemic agency. It is no different from treating someone like a tree from which we can infer information about its age by observing its rings. This implies that there are certain things to which epistemic subjects have a right in relation to others, such as receiving a fair credibility assessment. When this right is not fulfilled, one is treated as an object. When one suffers from epistemic objectification, one is not treated as fully human. This is the primary harm of *testimonial* injustice.

Might metalinguistic injustice be construed as a testimonial injustice? The presence of metalinguistic power might be due to a speaker affording less credibility to another who advocates for an alternative word-concept pair. For example, the woman who accuses the man of sexual harassment might not be taken seriously when she endorses a 'sexual harassment'-concept pair that conflicts with the pair preferred by the man. However, it is also possible that one can possess metalinguistic power owing to the operation of identity prejudice *without* a speaker being subject to a deflated credibility assessment. For instance, a marginalized speaker might have internalized pernicious stereotypes about their own reliability, and thus doubt themselves and their preferred word-concept pairs. Given this, the marginalized speaker might limit their speech in metalinguistic negotiation. Consequently, the hearer gains metalinguistic power. Yet, the testimony of the speaker isn't given an unfair

credibility assessment since little testimony was given.[161] Thus, metalinguistic injustice can occur without testimonial injustice. Metalinguistic injustice is *not* testimonial injustice.

### 6.4. Prejudice in the Hearer: Linguistic Objectification

The claim that the primary harm of testimonial injustice is epistemic objectification has its critics. José Medina argues that one can be treated as an informant and still be undermined as an epistemic subject, such as when an informant is not treated as an inquirer (2013, p. 203 – 204). Gaile Pohlhaus Jr. (2014) argues that it is unintuitive to think of the primary harm of epistemic injustice as a kind objectification since some forms of epistemic injustice structurally depend on the epistemic subjectivity of a speaker (Pohlhaus Jr. 2014, p. 103):

> When I count the rings on a felled tree, I cannot plausibly deny that it has the number of rings that I have just counted. When I see a women shaking rain off her umbrella, I cannot reasonably fail to believe that her umbrella is wet. Moreover, if I cannot bring myself to believe these things, others may rightly question my judgment and call me irrational. Furthermore, if I am not irrational, in the face of my inability to believe that the tree has the number of rings I have counted or that the umbrella is wet, I may wish to seek professional help concerning my cognitive state. Indeed, certain forms of agnosia present us with exactly the kind of case where there is no sensory impairment but a person cannot bring himself to believe what he ought to infer from his senses– for example, that this is his very own leg or that is his very own child. And yet, in cases of testimonial injustice, it is precisely this kind of inability to believe that becomes viable. Epistemic objects, in this sense, make a kind of claim on the knower that is denied to the victim of testimonial injustice (Pohlhaus Jr. 2014, p. 103).

Finally, Aiden McGlynn (2019) argues that the main problem with Fricker's account is that it seems that when one perpetrates epistemic injustice, they are not treating someone as lacking epistemic agency in the sense of a mere source of information.

Despite these concerns, McGlynn contends that Fricker's account of epistemic objectification is stronger than most critics assume. In an attempt to save Fricker's account, McGlynn suggests that we should look more closely at Nussbaum's (1995) seven ways of treating someone as an object (ibid. p. 12). I won't spend time listing them here. Nevertheless, McGlynn argues that if Fricker expands her notion of objectification she can accommodate the cases that have troubled her.

This insight helps to understand another plausible means of construing the wrong of metalinguistic injustice as linguistic injustice, which brings us to (1). Instead of metalinguistic injustice being a form of *epistemic* objectification, we can think of it as a form of *linguistic* objectification. What is linguistic objectification?

A first guess is that it is the denial of an agent's capacity to do certain things with words. One is linguistically objectified when one loses control over exercising basic linguistic capacities, such as describing one's experience, communicating it across social space, being the receiver of such information, performing certain actions with an utterance, etc. It isn't hard to recognize these capacities as fundamental and distinctively human (Hornsby and Langton 1998, p. 37). They enable speakers to convey information across social space, which underwrites coordination of thought, talk,

[161] This idea is similar to Kristie Dotson's (2011) notion of *testimonial smothering*.

and practice; all of which are paramount for one to flourish (ibid. p. 37). Thus, when one's linguistic capacities are restricted, one is undermined as a linguistic subject.

Understood in this sense, linguistic objectification comes close to Nussbaum's (1995) notion of *inertness,* one of her seven forms of objectification. The objectifier treats the subject as *lacking agency*; for my purposes, linguistic agency. This is on the right track, however it admits cases that seem unintuitive. A classroom rule of 'no swearing' prevents students from saying curse words. This looks like a non-problematic denial of agency. Thus, in order for linguistic objectification to constitute linguistic *injustice*, we might say that one has to lose control over their words *under certain conditions*. Linguistic objectification occurs whenever metalinguistic power is in play. That is, linguistic objectification occurs when one loses control of their words *owing to identity prejudice.* And there appear to be clear examples of this in existing literature.

Consider *silencing.* This refers to the systematic interference with communicative capabilities of a speaker in conversation (McGowan 2017, p. 39). It is generally agreed that problematic forms of silencing involve a causal connection between one's disadvantaged social identity and one's (in)ability to exercise basic linguistic capacities. Which capacities or capabilities are undermined in a context depends on the nature of the situation.

One might not have the capacity to perform certain speech acts. This is called *illocutionary disablement*. In such cases, a speaker utters words yet those words fail to constitute an intended action. Langton and Hornsby (1998) argue that this kind of silencing owes to a communicative interference that constitutes uptake failure. For example, in a social environment in which men are heavily influenced by pornography, women are often unable to perform the act of refusing sex since 'no' is taken up as 'yes' by men.[162] The wrong lies in the systematic failure of members of one identity group to recognize the illocutionary intentions of members of another identity group owing to being situated in a particular social environment.

Another way that linguistic capacities might be unfairly affected is through *perlocutionary frustration* (Langton 1993). This occurs when a speaker successfully pulls off a speech act, yet the speech act fails to have its intended effect on conversation. Continuing with our example, consider a woman who refuses sex and a man recognises her illocutionary intention yet fails to take seriously the implications of this act. The refusal is successful, but its intended effect is frustrated. She is denied an outcome in conversation to which she is entitled.[163] The wrong lies in the systematic failure of members of one identity group to respect the communicative implications of the speech of members of another identity group owing to being situated in a particular social environment.

We, as humans, are entitled to exercise certain linguistic capacities, whether illocutionary or perlocutionary. This constitutes (part of) our linguistic agency. When such capacities are undermined owing to the operation of identity prejudice, this constitutes linguistic objectification. And, linguistic objectification is the primary harm of linguistic injustice. We might ask: Does linguistic objectification occur in particular cases non-ideal metalinguistic negotiation?

Consider the exchange over the 'sexual harassment'-concept pair. As discussed, the man has metalinguistic power over the woman. Specifically, the woman cannot perform m-moves with same impression as the man, subsequently having less influence over the word-concept pair that will be

[162] For others who have similar, yet competing views about illocutionary silencing, see Maitra (2009), Kukla (2014), Hesni (2018).

[163] McGowan (2017) identifies three (additional) ways in which one might suffer from this kind of communicative interference in refusal cases. A hearer might fail to recognize: (i) the authority of a speaker, (ii) that the speaker is sincere, or (iii) the speaker's 'true feelings' (2017, p. 47 – 50). And there may be further possibilities too.

operative in the context. The woman cannot (fully) exercise particular metalinguistic capacities. She cannot leave her desired mark on conversation. Put differently, she suffers from a form of perlocutionary frustration at the level of metalinguistic discourse. Further, she cannot exercise such capacities *in virtue of being a woman.* Identity prejudice plays a role in what the woman can do with her words in metalinguistic negotiation. Thus, she suffers from linguistic objectification – the operation of metalinguistic power involves a linguistic injustice. There is a causal connection between the speaker's disadvantaged social identity and her (in)ability to exercise certain metalinguistic capacities. Call this *metalinguistic injustice (LI*).*

> **Metalinguistic Injustice (LI*)** occurs when a speaker suffers from linguistic objectification owing to the operation of metalinguistic power in a context.

Metalinguistic injustice (LI*) is not simply preventing someone from performing certain speech acts – it is not just illocutionary disablement. Metalinguistic injustice (LI*) includes cases in which one can perform a desired speech act, but nevertheless fail to leave a strong impression with the performance of this act. Metalinguistic Injustice (LI*) is often a form of perlocutionary frustration.

A final note about metalinguistic injustice (LI*). Unlike metalinguistic (EI) and (LI), which depend on the aim of completing metalinguistic negotiation in order to determine the epistemic or linguistic resources of a context, metalinguistic injustice (LI*) can occur without this aim. It simply involves one suffering linguistic objectification at the metalinguistic level, whether or not the goal is to settle on the concept that will be expressed by a word in a context.

I want to end this section by saying that metalinguistic injustice could, and perhaps does, involve all that has been suggested. Metalinguistic injustice is either metalinguistic injustice (EI), (LI), (LI*), or all simultaneously.

### 6.5. Comparisons

I want to spend time distinguishing my view from closely related injustice: *contributory injustice* and *contextual injustice.* I'll start with the former.

Kristie Dotson (2012) argues that many cognitive resources are developed from marginalized situatedness. And often, such resources are not shared with members of dominant out-groups. Thus, marginalized folk are often in a position to understand their own experience and are able to describe such experience to members of the same group, but cannot communicate their understanding to dominant out-group members. She calls this *contributory injustice.* The essence of the injustice is that marginalized folk are unable to contribute equally to the collective hermeneutical resource of a community of thinkers and speakers in virtue of their contributions being systematically dismissed owing to willful ignorance connected to structural identity prejudice.[164]

The similarity between contributory injustice and metalinguistic injustice (EI) and metalinguistic injustice (LI) is reasonably easy to see. Namely, all three involve the possession of a resource by a marginalized person that is dismissed as a contribution to shared understanding by a dominantly situated speaker. Despite this similarity, there are clear differences. The first, and obvious, is that metalinguistic injustice (LI) concerns one's inability to contribute the set of shared *linguistic* resources. However, this hardly seems worth mentioning. The second, and still obvious, difference is that contributory injustice concerns the inability of marginalized folk to contribute to the *collective*

[164] This is similar to Pohlhaus's (2012) notion of willful hermeneutical ignorance.

resources of an epistemic/linguistic community. Metalinguistic injustice concerns the wrong that is involved in *interpersonal communication*. This difference makes explicit the connection between metalinguistic injustice and contributory injustice: Metalinguistic injustice is one mechanism by which contributory injustice occurs. It is in virtue of the fact that a good deal of metalinguistic disagreements are resolved in favour of dominantly situated speakers that word-concept pairs developed from marginalized situatedness aren't able to reach the level of public consciousness. After all, if the word-concept pairs developed from marginalized situatedness aren't able to even get airtime in conversation, they stand little chance of reaching collective understanding.

The second closely related injustice that I will discuss is Jonathan Jenkins Ichikawa's (2020) notion of *contextual injustice*. This occurs when disadvantageous contextual parameters are set during conversation. Ichikawa's primary interest is in the unjust setting of disadvantageous standards for knowledge. However, he takes the injustice to generalize to other context-sensitive expressions. Here is an example. When assessing the testimony describing sexual harassment, one might employ unnecessarily high standards for knowledge by raising pernicious error possibilities (e.g., 'Are you sure it wasn't an accident?'). For Ichikawa, the wrong of contextual injustice is closely linked to Fricker's notion of testimonial injustice (2020, p. 14). It involves a situation in which a speaker's testimony is undermined in virtue of denying it the status of 'knowledge.'

There are a couple of ways that metalinguistic injustice contrasts with contextual injustice. The first is that contextual injustice can occur with the use of only first order utterances. A speaker needn't engage in metalinguistic disagreement in order to set unfair standards for context-sensitive expressions – they simply have to raise error possibilities. For example, contextual injustice can occur without engaging in a back-and-forth about the concept that should be expressed by the term 'sexual harassment.' One simply has to say, 'Are you sure it wasn't an accident?' This, however, ignores the (unequal) power at play in setting standards through metalinguistic discourse.

This connects to a second, and important, way that metalinguistic injustice contrasts with contextual injustice. Ichikawa locates the wrong of contextual injustice in the *setting of unfair standards that are disadvantageous for a marginalized speaker.* I locate the wrong of metalinguistic injustice prior to this. Metalinguistic injustice (EI) and (LI) involve the denial of someone from contributing to the epistemic/linguistic resources of a local context; and metalinguistic injustice (LI*) involves denying someone the ability to exercise linguistic agency – exercising certain (meta)linguistic capacities. What's particularly important about this is that metalinguistic injustice (EI), (LI), and (LI*) often sets up the conditions under which contextual injustice occurs. Unfair standards are set in a context *in virtue of* the operation of metalinguistic power that governs how concepts are paired with terms. Thus, metalinguistic injustice *precedes* contextual injustice.

## 6.6. Strategies for Resistance

I will not discuss how metalinguistic power and deprivation relate to conceptual reproduction and the implementation challenge of conceptual engineering. This was explored at the end of the foregoing chapter. In this section, then, I will focus on something else. Given that we have located a specific wrong that occurs in metalinguistic negotiations that have illegitimate power running through them, this puts us in a position to offer corrective measures. I will not be able to offer a comprehensive account, but I will gesture at some possibilities. Such possibilities are restricted to those who are motivated to avoid perpetrating moral wrongs. My hope is that this will go some way to helping us think about how we can work against the threat of accidentally reproducing ideological concepts.

In cases of hermeneutical injustice, a disadvantaged speaker is unable to render communicatively intelligible certain experiences owing to a gap in the collective hermeneutical resource. Thus, to overcome this, a dominant hearer must take on a normative burden. But exactly what can a dominant hearer do to correct hermeneutical injustice in specific instances? Fricker suggests,

> ...the form the virtue of hermeneutical justice must take, then, is an alertness or sensitivity to the possibility of that the difficulty one's interlocutor is having as she tries to render something communicatively intelligible is due not its being nonsense or her being a food, but rather to some sort of gap in collective hermeneutical resources. The point is to realize that the speaker is struggling with an objective difficulty and not a subject failing (2007, p. 169).

What does this sensitivity look like? According to Fricker, the virtuous hearer must cultivate a *reflexive critical openness;* one must be aware that a marginalized speaker is having trouble articulating their experience, and how this connects to systems of privilege and subordination. The aim is to 'neutralize the impact of structural identity prejudice on one's credibility judgement' (Fricker 2007, p. 173) in order to create a more inclusive hermeneutical climate.

I think the virtue approach is useful, though this particular strategy doesn't appear to be a good one for overcoming metalinguistic injustice. This is for two reasons. The first is that a person who has metalinguistic power may be able to perfectly understand what the person suffering from metalinguistic deprivation is saying. Thus, there is no concern about intelligibility. Further, it isn't enough that the person who has metalinguistic power be a receptive hearer – after all, one can be a receptive hearer yet still disagree with someone on the basis of conceptual difference.

Despite this, there is a sense in which there is an issue of hermeneutical sensibility. It is an issue that concerns whether one has entertained the question of whether the concept that they are deploying is better suited for the context than the concept used by an interlocutor. When one takes time to critically question their own concepts relative to an interlocutor, in a way that is as free from bias as possible, call this the virtue of *hermeneutical humility*. Importantly, this virtue appears to be a viable solution to some instances of metalinguistic injustice: it significantly reduces the chances of someone suffering from metalinguistic deprivation insofar as it requires that one take seriously another's conceptual understanding.

But, as Rae Langton (2010) asks of Fricker: Is virtue enough?[165] Individual or agential remedies might be a good starting point, but surely we must pay attention to social structure? After all, as Pohlhaus Jr. suggests, using the wrong concepts can involve a 'systematic and coordinated misinterpretation of the world' (2012, p. 731). Perhaps we should focus our energy on the socio-epistemic processes or structures involved in determining the concepts we collectively have access to which will have a substantive bearing on the concepts we choose to use in interpersonal settings. That is, perhaps we should be less focused on bottom-up solutions to metalinguistic injustice, and instead focus on top-down ones. In saying this, we do not have to choose one solution. It seems entirely plausible that our collective resources, epistemic or linguistic, are influenced by the everyday decisions we make about the concepts we choose to use in conversational contexts. However, it also seems plausible that the everyday decisions we make about which concepts to use in conversation are influenced by our access to collective resources. Thus, I think that hermeneutical humility is

[165] Others have objected to Fricker's virtue solution on different grounds. See Alcoff (2010) and Sherman (2016).

something that we should individually strive for, while at the same time being motivated to take steps toward a conceptual revolution.[166]

## 6.7. Conclusion

To sum up, in this chapter I have explicated the notion of metalinguistic injustice. It is an injustice that occurs in metalinguistic disagreement in which illegitimate power is present – power that is derived from the operation of identity prejudice that constrains what another speaker can achieve in the joint activity of pairing contents with words. I argued that we can understand metalinguistic injustice in a number of ways, broadly divided into camps that spell out the injustice as either identity prejudicial exclusion from the pooling of resources, or identity prejudice in the hearer. With respect to the former, I first argued that metalinguistic injustice constitutes an epistemic injustice insofar as one can be excluded, on the basis of identity prejudice, from the pooling of local cognitive resources. I called this metalinguistic injustice (EI). I then argued that metalinguistic injustice can constitute a linguistic injustice insofar as one can be excluded, on the basis of identity prejudice, from the pooling of local linguistic resources. I called this metalinguistic injustice (LI). And with respect to the latter, I argued that metalinguistic injustice constitutes a further linguistic injustice insofar as a speaker suffers from linguistic objectification owing to identity-prejudicial metalinguistic deprivation in a context.

In the next chapter, I continue to discuss metalinguistic injustice and its associated wrongs. In particular, I examine how metalinguistic injustice can help us to understand a distinctive form of gaslighting. This form of gaslighting occurs when there is disagreement over whether a shared concept applies to some aspect of the world, and where the use of words by a speaker is apt to cause a hearer to doubt her interpretive abilities *without* doubting the accuracy of her concept.

[166] Nevertheless, I maintain that the primary focus of the implementation challenge should be on revising our socio-epistemic structures.

# Chapter Seven

## Gaslighting, First and Second Order

*Individuals engage in... gaslighting when they invoke the epistemologies and ideologies of domination that actively disappear and obscure the actual causes, mechanisms, and effects of oppression.*

Nora Berenstain 2020, p. 734

### 7. Introduction

This chapter continues the discussion on metalinguistic injustice that was explored in the previous chapter. Specifically, I examine how metalinguistic injustice helps us to locate a distinctive form of *gaslighting*. I draw a distinction between *first order* and *second order* gaslighting, where my primary interest is in the latter. First order gaslighting occurs when there is disagreement over whether a shared concept applies to some aspect of the world, and where the use of words by a speaker is apt to cause a hearer to doubt her interpretive abilities *without* doubting the accuracy of her concept. Second order gaslighting occurs when there is disagreement over which concept should be used in a context, and where the use of words by a speaker is apt to cause a hearer to doubt her interpretive abilities *in virtue of* doubting the accuracy of her concept. I suggest that many cases of second order gaslighting are *unintentional*: its occurrence often depends on contingent environmental facts. Second order gaslighting is a distinctive mechanism by which ideological concepts are informally and locally reproduced (and how the proposal of alternative, and better, concepts can be undermined).

### 7.1. What is Gaslighting?

Gaslighting is a contested notion.[167] To get clear on how I'll make sense of it, I want to start with ordinary understanding. A quick online search of gaslighting returns a common theme. This theme is best captured in Stephanie Sarkis's (2017) *Psychology Today* article where she writes:

> Gaslighting is a tactic in which a person or entity, in order to gain more power, makes a victim question their reality (2017, p. 1).

Roughly, this definition suggests that gaslighting is particular behaviour by one party that is causally responsible for making another come to doubt their understanding. This is vague, but it serves as a useful starting point.

The first thing to notice is that this definition makes a strong causal claim. According to Sarkis, gaslighting tactics *make* a victim question their reality. However, I want to accommodate cases where such tactics are unsuccessful. Thus, I take gaslighting to involve behaviour that is *apt* to make a victim question their reality.

[167] For competing accounts, see Abramson (2014), Spear (2018, forthcoming), Stark (2019).

To fill out more of the details, we should ask: What kind of behaviour is gaslighting? I will treat gaslighting as a *linguistic* phenomenon. I accept that it could encompass much more than exchanging words, but it is not my aim to shed light on such possibilities. My interest is in forms of gaslighting that involve language use. Specifically, the use of language *apt for inducing certain attitudes in a hearer*. Which attitudes? Those that concern one's *interpretive abilities.* This includes: the ability to get facts right, and the ability to properly evaluate situations (Abramson 2014, 8). Thus, to say that gaslighting involves linguistic behaviour that is apt to cause someone to 'question their understanding of reality' is to say that such behaviour is apt to cause someone to doubt their interpretive abilities.

What this *doesn't* say is that gaslighting merely involves getting someone to believe that their understanding is mistaken. Gaslighting affects *higher-order* attitudes. For someone to be gaslit, they must form *negative attitudes about their attitude-forming mechanisms.* In other words, gaslighting affects one's stance towards one's own reliability. Consider a familiar case of gaslighting. Imagine a wife who sees her husband kiss another woman. When she confronts the husband about this, he says that he was simply greeting the other woman and that she is 'crazy' for thinking otherwise. Here, the husband is not simply telling his wife that she is mistaken. He is saying something about the *means* by which she is mistaken. He is asserting that her ability to form appropriate attitudes about the situation is unreliable. For gaslighting to be successful, the wife must come to doubt her interpretive abilities, not merely her understanding of the event.

We can put this point another way: Gaslighting targets intellectual self-trust.[168] And intellectual self-trust is comprised of *cognitive* and *affective* elements (Jones 2012). A person who is subject to gaslighting might come to form a *belief*, or make a *judgment*, about the unreliability of their interpretive abilities (e.g., forming the belief that they are 'crazy').[169] However, cognitive elements do not account for all failures of self-trust. Self-trust can also be undermined by *affective attitudes*. After all, the doubt in one's own interpretive abilities may not be a response to reasons. Instead, it is due to certain emotions: anxiety, depression, hopelessness, lack of confidence, the need to come to a conclusion, etc. (2012, p. 6). To see this, consider the example above. When the wife confronts the husband after seeing him kiss another woman, she has sufficient reason to believe that her husband is cheating on her. However, let's say that owing to the power operating in the marriage, and the costliness of leaving, imagine that the wife still doubts what she saw. This needn't be because she believes that her perceptual and cognitive faculties have failed her. The niggling feeling of doubt is not a response to reasons. Instead, it is an affective stance toward her own interpretive abilities: The wife has her attention on the *bare* possibility that she might be mistaken.[170]

With this in mind, a first stab at defining gaslighting might run as follows:

---

[168] For detailed discussion on the relationship between gaslighting and self-trust see Spear (2018, forthcoming).

[169] In a domain, or across domains.

[170] Karen Jones (2012) has a really nice example that captures the role of affect in self-trust. She recounts an experience where she's at the airport constantly checking to see whether her passport is still in her pocket. She keeps checking her pocket even though she has no reason to doubt her perceptual and cognitive faculties. The feeling of doubt and anxiety is enough for her to disregard compelling reasons. To me, this shows that stakes are an important factor in affect and self-trust – after all, losing one's passport is a real nuisance when travelling in a foreign country. When it comes to gaslighting, stakes might also be an important factor. For example, if a wife catches her husband cheating, she might be more disposed to doubt herself if it means having to end the relationship. Given the power structures in play, leaving a marriage can be costly. Unfortunately I don't have room to explore this thought further.

**Naïve Gaslighting:** Gaslighting occurs when (i) a speaker uses words (ii) that are apt to cause a hearer to doubt her own interpretive abilities.

It is easy to produce counterexamples to *Naïve Gaslighting*. After all, it makes it seem that gaslighting is pervasive. I think that it is, but not as pervasive as this definition allows. It allows for situations in which someone quite innocently uses words that are causally responsible for another coming to doubt their interpretive abilities. Imagine a couple that disagrees on what the total cost of dinner will be, and the one who is wrong comes to form the belief that they are cognitively defective because of this. It is far too much of a stretch to say that this falls within the proper bounds of gaslighting.[171] Thus, *Naïve Gaslighting* doesn't specify the relevant conditions to be a definition of gaslighting.[172] So, there is a residual question: In order for speech to count as gaslighting, what is the relevant relationship between the use of words by a speaker and the inducement of certain higher-order attitudes in a hearer?

A simple answer to this is that an offender of gaslighting has to *intend* to induce doubt in a hearer. We can call this *intentional gaslighting*:

**Intentional Gaslighting:** Gaslighting occurs when (i) a speaker uses words with (ii) the intention that a hearer comes to form (iii) negative attitudes toward her own interpretive abilities.[173]

All that's important to notice about *Intentional Gaslighting* is that it requires that one intend to gaslight. Take the case above where the wife sees her husband kiss another woman. According to this definition, for the husband to count as gaslighting his wife, he must intend for her to doubt her interpretive abilities. This is perhaps what we have in mind when we think of paradigm cases of gaslighting.

One might take exception to *Intentional Gaslighting* on the grounds that not all intuitive cases of gaslighting involve an intention of this sort. Think of a situation in which a man brushes up against a woman's butt in the office, and the woman reports this to a colleague who responds, 'I'm sure it was innocent, John isn't the kind of guy to act inappropriately at work'. As a result, the woman comes to doubt her ability to recognize sexual harassment. In this case, the colleague may not intend for the woman to doubt her interpretive abilities, even though he expressed doubt in her testimony. Nevertheless, this looks like a cut and dry case of gaslighting.[174] The lesson seems to be that not all cases of gaslighting are intentional, though perhaps a great many of them are.

If we can't appeal to intentions to account for all cases of gaslighting, it seems that we have to look beyond psychological facts. A useful starting point is to recognize that the world is unjust. As we know from chapters five and six, certain groups of people are privileged, others are subordinated. One sense in which the world is unjust is that certain people are excluded from, or lack full

[171] And it would rob the notion of gaslighting of political usefulness.

[172] Note that 'naïve' doesn't refer to the gaslighter, but to the definition.

[173] We might want to add in a fourth condition that states that the intention to induce doubt in a hearer is *unwarranted*. This means that cases in which someone has good reason to convince another to doubt their interpretive abilities aren't counted as gaslighting. For example, one might get fed up with a conspiracy theorist who endorses the view that vaccines cause autism, and instead of offering arguments, one might simply tell the other person that they are 'crazy.' I want to keep open the question of whether this counts as gaslighting, though I admit it strays from typical cases.

[174] This case of unintentional gaslighting is similar to a case discussed by Rachel McKinnon (2017) in which a trans woman is misgendered by an ally, and when she reports this to a friend, the friend says that she must have misheard the ally.

participation in, the economy of epistemic relations; the social network in which we exercise our capacity as epistemic subjects. Exclusion from (full) participation in such relations is called *epistemic injustice* (Fricker 2007). This is a wrong that occurs when someone is undermined in their capacity as an epistemic subject. There are many ways that one might suffer from epistemic injustice. One might receive an unfair credibility assessment, and thus suffer from *testimonial injustice* (Fricker 2007, Ch. 1); one might be unable to render intelligible experience, and thus suffer from *hermeneutical injustice* (2007, Ch. 7); and/or one might be subject to ill-fitting interpretations, and thus suffer from *contributory injustice* (Dotson 2012) or *wilful hermeneutical ignorance* (Pohlhaus Jr. 2012), etc. Whatever the case, the important point to focus on is that identity-prejudice plays a substantive role in determining who is afflicted most by epistemic exclusion. Such injustice exists in our structures, and dictates the character of our epistemic lives.

If gaslighting affects one's attitudes toward their interpretive abilities, then it is easy to see why epistemic injustice is worth bringing up. Given that certain groups are excluded from, or lack full participation in, the economy of epistemic relations, then this serves as partial explanation for why members of such groups might be inclined to downgrade their self-trust in the face of doubt. After all, they have never really been treated as full epistemic subjects. Because of this, we might think that an unintentional story of gaslighting should be sensitive to such facts. Thus, we can define unintentional gaslighting as follows:

> **Unintentional Gaslighting:** Gaslighting occurs when, (i) a speaker uses words *without* the intention that a hearer come to form negative attitudes toward her own interpretive abilities, but (ii) the use of such words is apt to cause the hearer to form such attitudes (iii) owing to the hearer being subject to systematic epistemic injustice that has disposed her to do so.

*Unintentional Gaslighting* allows us to distinguish between cases in which it seems that gaslighting has occurred and cases in which it seems that it hasn't, but where the only relevant difference is that one person is subject to systematic epistemic injustice, and the other is not.

For example, consider the case above where the woman testifies to being deliberately brushed up against by a man in the office, and her colleague responds, 'John isn't the kind of guy to act inappropriately at work'. In a society that systematically fails to treat the testimony of women seriously when it comes to allegations of sexual harassment (and sexual assault), gaslighting is more of a possibility. Why? The tendency of members of society to disbelieve women when such allegations are made disposes women to doubt their own experience involving sexual harassment. And this doubt is not just about *particular* cases of sexual harassment, but the ability to identify it *generally*. Women are pushed to doubt their interpretive abilities (in a domain, or across domains). They are told: 'nice guys don't sexually harass people;' 'you asked for it by wearing provocative clothing;' 'just accept that's what men are like;' etc.

Compare this case with an otherwise identical one, but where it is a man who gets brushed up against and is told 'John isn't the kind of guy to act inappropriately at work'. On the unintentional reading, this doesn't count as gaslighting. I think that this gets the right result. This is not to say that sexual harassment didn't occur. It is only to say that when the man's allegation of sexual harassment is dismissed, it isn't a case of gaslighting. Why? Men enjoy a great deal of credibility. They are not disposed to doubt themselves in the face of conflicting understanding. And this is due to the fact that men don't suffer from pernicious stereotypes concerning their ability to get facts right or properly

evaluate situations (e.g., men are not associated with being too emotional or sensitive).[175] The important difference between the two cases is this: Women are subject to systematic credibility deflations and hermeneutical deficiencies, whereas men are not. Unintentional gaslighting is sensitive to such facts.

Given the foregoing considerations, a definition of gaslighting should accommodate cases that are intentional and unintentional. That is, intentional and unintentional gaslighting are complements – both are relevant to gaslighting. Thus, I suggest the following *disjunctive* definition:

> **Gaslighting:** Gaslighting occurs when (i) a speaker uses words and either (ii) the speaker intends for the use of such words to cause a hearer to form (iii) negative attitudes toward her own interpretive abilities, *or* (iv) a speaker uses words without such an intention, but (v) the use of words is apt to cause the hearer to doubt her interpretive abilities (vi) owing to the hearer being subject to systematic epistemic injustice that has disposed her to do so.

With this definition in mind, I want to discuss one further aspect of gaslighting that is paramount. Bringing to light the nature of unintentional gaslighting exposes its *temporal* aspect. We might tend to think of gaslighting as one-off events, but it is perhaps better thought of as something that occurs within a system of oppressive patterned behaviour *over time*. We can explain why it is that people are more disposed to doubt themselves in the face of certain utterances by looking at the social environment in which they are situated, and the information that is available in such environments that enable coordination between members.

As discussed in chapter five, a social environment is a network of interrelated and regular patterns of coordinated social behaviour. Such patterns depend on the culturally available information that agents draw on, such as coordinated concepts. This information structure frames expectation, provides action-options, and serves to stabilize behaviour by providing quasi-rules that govern social interaction. Agents rely on such information to render intelligible experience, which then facilitates intentional engagement and coordination with others and surroundings. Moreover, social environments can be unjust. This occurs when the extant patterns of coordinated behaviour privilege some, and subordinate others in unjust ways.

To coordinate at a given time, social agents draw on certain concepts and characterizations that narrow the range of expected behaviour. The salience of such concepts can be made explicit by the interacting parties, or the context in which the parties are situated. For example: In an oppressive patriarchal society, pernicious patterns of interaction are activated when an ideological concept of woman is salient in certain domains. A pertinent case is the systematic bias in assessment of intellectual contributions by women in philosophy. Here, it is the context of philosophy that makes gender a salient cultural tool to use when evaluating intellectual credibility. Other contexts include male dominated workplaces and domestic settings.

Much of gaslighting, intentional and unintentional, depends on the salience of (social) concepts; classificatory practices and characterizations. This is because concepts determine action-options and thus undergirds behaviour. Consider the example above where the woman testifies to being sexually harassed, and the colleague responds by saying 'John isn't the kind of guy to act inappropriately at work.' Despite the word 'woman' not being explicitly mentioned in the interaction, an associated concept is made salient given the nature of the conversation, and the context in which the

[175] Even if the man did come to doubt himself after being dismissed, it would still not count as gaslighting. This is because the man is not subject to systematic credibility deflation due to him *being a man.*

conversation takes place. After all, a *woman* has made an accusation against a *man* in a male dominated workplace situated in a patriarchal social environment. Due to the association between women and characterizations such as 'hysteria,' 'over-sensitivity,' and 'paranoia' in this environment, what is fitting for the woman to do in the face of doubt is to downgrade her self-trust. In other words, women are disposed to downgrade their self-trust given persistent challenges to their epistemic reliability. Thus, the success of gaslighting owes much to the salience of certain concepts (e.g., identity-prejudicial stereotypes and forms of classification) that govern social interaction.

In sum, gaslighting is not something that usually happens in one-off, isolated instances. Instead, it is often a function of the social environment in which one is situated that has an extended history of unjust treatment of (some of) its members. The success of gaslighting is dependent on the background patterns and culturally available information that gaslighters are able to tap into.

### 7.2. Different Cases of Gaslighting

Up to this point, I have only discussed how I'll make sense of gaslighting. It is a linguistic phenomenon that involves the use of words apt for causing a hearer to doubt her interpretive abilities. This can be intentional or unintentional, and often requires making salient pernicious cultural tools (e.g., stereotypes). Now I want to explore the following question? *In what sense does gaslighting cause someone to doubt their interpretive abilities?* What I will suggest is that there are two broad ways of answering this question. Each answer depends on the linguistic order at which a gaslighting exchange takes place. *First order* gaslighting occurs when a gaslighter and the subject of gaslighting[176] aim to coordinate on a presupposed shared concept, but take themselves to disagree as to whether the concept applies to some aspect of the world. *Second order* gaslighting occurs when a gaslighter and the subject of gaslighting *do not* aim to coordinate on a presupposed shared concept and, instead, disagree as to which concept should be used in a context.[177] Second order gaslighting has been overlooked as a distinctive form of gaslighting in the philosophical literature. This may be due to the fact that very many instances of second order gaslighting seem unintentional – it is not the sort of thing that comes to mind when we think about clear cases of gaslighting. Nevertheless, it is a form of injustice that should be made explicit.

We can see gaslighting as occurring at different linguistic orders by comparing the following pair of conversations:

(1) Woman: John brushed up against my butt; that's sexual harassment.
Man: Sexual harassment? I'm sure it was an accident.

(2) Woman: John brushed up against my butt; that's sexual harassment.
Man: That's not sexual harassment. It's so trivial.

[176] The phrase 'subject of gaslighting' is a mouthful. Others in the literature have used phrases such as 'target' or 'victim' of gaslighting. I disagree with the former on the grounds that it gives the impression that gaslighting is always intentional. And I wish to avoid the latter because the language of victimhood has negative social meaning attached to it.

[177] This is not to be confused with the claim that the gaslighter and the subject of gaslighting aren't trying to coordinate on *some* concept – this I take to be what we're are doing all the time. The claim is that the gaslighter and the subject of gaslighting aren't trying to coordinate on either of their preferred concept.

Both conversations represent a fairly typical gaslighting exchange: Someone offers testimony, and this testimony is rejected out of hand.[178] So, what we can see is *disagreement.* By 'disagreement,' I don't mean a rational dispute between epistemic peers – there is no reasons-giving dialogue. Instead, I mean only to say that there is a difference in what the woman and man take to be the nature of the event that took place. In both (1) and (2), the woman thinks that John brushing up against her is sexual harassment, and the man does not.

What we can also see is that the *grounds* on which the man rejects the woman's understanding of the event as sexual harassment differs between (1) and (2). In (1), the man objects to the woman on the grounds that it must have been an accident. Perhaps this is due to his affection for John, or his general skepticism of workplace harassment. In (2), the man objects to the woman on the grounds that she doesn't understand what sexual harassment is. The man has a different idea of the range of things that count as sexual harassment. In short: (2) involves a presupposed *conceptual difference*; (1) does not.

## 7.3. Orders of Usage and Disagreement

Speakers use words to convey information about the world. When I say 'Fil is tall,' I am describing how things are: There is someone named Fil, and he has the property of being tall. However, as was discussed in chapters four and five, sometimes words are used by speakers to convey information about the appropriate or correct usage of those words in a context. In such cases, when I say 'Fil is tall' I am pragmatically advocating for how to use 'tall' in the context. Call the former way of using words *first order usage*, and, following chapter five, call the latter *metalinguistic usage*.

Disagreements can occur at both orders of usage. I can say 'Fil is tall,' and someone can disagree with me by saying 'Fil is not tall'. What ensues is a disagreement that centers on whether or not a threshold of tallness has been met. We can call this *first order disagreement.* First order disagreements concern what the world is like, where speakers are committed to coordinating on a presupposed shared concept – e.g., speakers agree on a threshold that must be met for being tall, but disagree as to whether someone has met this threshold. However, disagreements can also occur over how 'tall' *should* be used in a context. Speakers can disagree as to which threshold must be met in order for someone to be tall. We have been calling this *metalinguistic negotiation* – we might also call it *second order disagreement.*

## 7.4. First Order Gaslighting

With this distinction in mind, let's consider (1) once more:

(1) Woman: John brushed up against my butt; that's sexual harassment.
Man: Sexual harassment? I'm sure it was an accident.

Here, it appears that the woman and man are aiming to coordinate on a shared concept of *sexual harassment.* This is evidenced by the fact that the man contests whether the act was *deliberate,* and does not contest whether the deliberate act *counts* as sexual harassment. That is, the woman and the

[178] Gaslighting can occur without the subject of gaslighting offering testimony. For instance, someone might be subject to constant claims that they are 'crazy' throughout a relationship, even without saying a word, and this can prompt the recipient of such claims to doubt their interpretive abilities.

man both agree that *if* the act of brushing up against the woman's butt were deliberate, it would have been sexual harassment. There is no presupposed conceptual difference. Thus, (1) is first order disagreement.[179]

How does first order disagreement turn into first order gaslighting? It depends on whether the man intends to gaslight. On the intentional reading, the man recognizes that the woman has a competing interpretation of the event, and with this recognition, he intends to cause the woman to doubt her ability to make reliable judgments about whether an event falls within the range of their shared concept of *sexual harassment.*

On the unintentional reading, the man recognizes that the woman has a competing interpretation of the event, however this recognition does not undergird an explicit intention of getting the woman to doubt her interpretive abilities. Instead, the man simply takes the woman to be wrong about the bare facts of the situation. He dismisses her testimony. However, as suggested before, the context of this dismissal matters when it comes to gaslighting: Given the content of the conversation, the salience of certain cultural tools causes the woman downgrade her self-trust in the face of doubt. That is, the woman is part of a society in which allegations of sexual harassment are regularly dismissed, and this disposes her to doubt her interpretive abilities when her testimony is met with disbelief. Despite the man not intending to gaslight, he counts as gaslighting owing to contingent environmental facts.

With this in mind, we can define first order gaslighting as:

> **First order gaslighting** occurs only if (i) a gaslighter and the subject of gaslighting aim to coordinate on a presupposed shared concept, but where (ii) the gaslighter's use of words is apt to cause the subject of gaslighting to doubt her interpretive abilities, and (iii) this is *not* due to the subject of gaslighting doubting the accuracy of her concept.[180]

So, in what sense does someone doubt their interpretive abilities when subject to first order gaslighting? According to this definition: the subject to gaslighting doubts their ability to get facts right, or properly evaluate situations, but this doubt is not a result of mistrusting the accuracy of the concept used to pick things out in the world.

### 7.5. Second Order Gaslighting

Consider (2) once more:

(2) Woman: John brushed up against my butt, that's sexual harassment.[181]
Man: That's not sexual harassment. It's so trivial.

[179] It needs to be made clear that the man believes that the event is *not* sexual harassment, and not that it is sexual harassment but there are excusing factors.

[180] Notice that this definition does not specify sufficient conditions.

[181] I want to highlight the fact that 'sexual harassment' has a legal definition. Because of this, there are a couple of ways that we can understand the second order dispute. First, we can understand the dispute as the participants operating within the bounds of the legal definition of 'sexual harassment,' but contesting some of its constituent terms. Second, we can understand dispute as concerning the legitimacy of the legal concept itself – the legal definition may not be just or accurate (or both). Thank you to an anonymous reviewer for this suggestion.

Here, it appears that the woman and man are not aiming to coordinate on a presupposed shared concept of *sexual harassment*. They appear to have different concepts. This isn't to say that the woman and man are not aiming to coordinate on *some* concept. In a conversation where speakers recognize that there is conceptual difference, it stands to reason that they will do as much as possible to ensure that there is conceptual similarity.[182] After all, without establishing conceptual similarity, conversation is difficult to advance – either speakers will be talking past each other, or at least one speaker will not understand what is being said. What I am claiming is simply that the man and woman did not enter the conversation with the same concept. There is presupposed conceptual difference. Nevertheless, there is also an assumed background aim of coordinating on a concept that they can both agree on.

Not only are the man and woman failing to coordinate on a presupposed shared concept, they are also disputing which concept should be expressed by 'sexual harassment'. Why believe this? We can appeal to Amie Thomasson's (2017) diagnostic markers. On (i), the man and woman both accept the fact that John brushed up against the woman. Unlike (1), the man does not contest whether the act of brushing up against the woman's butt was deliberate. He accepts that it is deliberate, but still fails to count as sexual harassment. Thus, the disagreement is over how this deliberate act ought to be conceptualized. On (ii), we can imagine the exchange continuing where appeals are made to previous or actual linguistic usage, but where such appeals do not settle the disagreement. The man might say to the woman, 'sexual harassment is about actually groping women, what you experienced was just harmless flirting,' and the woman might respond, 'that's not how people use the term anymore'. On (iii), we can imagine that the man and woman realize that they are using the term 'sexual harassment' in different ways, but nevertheless they continue to disagree. The woman thinks the term should be used to include cases in which one is deliberately brushed up against; the man thinks that such cases are far too trivial to count. From this, we can see is that (2) meets the diagnostic criteria for second order disagreement.

Now, how does second order disagreement turn into second order gaslighting? We can tell roughly the same story as we did about first order gaslighting. The primary difference is *why* the subject of gaslighting comes to doubt her interpretive abilities. In first order gaslighting, the gaslighter's use of words is apt to cause the subject of gaslighting to doubt her ability to recognize sexually harassment *without* her questioning the accuracy of her concept of *sexual harassment*. That is, the subject of gaslighting comes to doubt her ability to recognize sexual harassment, but not because she believes that she has the wrong concept to identify it. Contrast this to second order gaslighting. The gaslighter's use of words is apt to cause the subject of gaslighting to form negative attitudes about her ability to recognize sexual harassment *in virtue of doubting the accuracy of her concept.* In other words, the subject of gaslighting comes to doubt her ability to judge whether sexual harassment has occurred *because she doubts that she has the right concept to pick it out.*[183]

With this in mind, we can define second order gaslighting as:

[182] This thought is expressed by Mitchell Green (2016) who claims that when speakers recognize that contexts are defective (when one speaker is mistaken about what is common ground), they will do what they can to correct it.

[183] Note that the gaslighter and subject of gaslighting needn't take themselves to be having a disagreement about concepts. This awareness is too cognitively demanding. In a sense, *recognition* that the dispute is about concepts may be a first step for countermeasures against gaslighting. Thank you to an anonymous reviewer for raising the point.

**Second order gaslighting** occurs only if, (i) a gaslighter and the subject of gaslighting *do not* aim to coordinate on a presupposed shared concept, and where (ii) the gaslighter's use of words is apt to cause the subject of gaslighting to doubt her interpretive abilities, (iii) *owing* to the subject of gaslighting doubting the accuracy of her concept.

We can ask: In what sense does someone doubt their interpretive abilities when subject to second order gaslighting? The answer: The subject of gaslighting doubts her ability to get facts right, and properly evaluate situations, in virtue of doubting the accuracy of the concept used to interpret some aspect of the world. It strikes me that much of second order gaslighting is unintentional.[184] Of course, this claim needs empirical backing. However, it seems that the second order gaslighter does not explicitly intend to cause the subject of gaslighting to doubt her interpretive abilities *through* challenging concepts. To accept otherwise would make gaslighting far too cognitively demanding, meaning that we would miss out on explaining many of its occurrences. Despite this, we might say that when a gaslighter challenges a concept, the challenge *implicitly* carries the assumption that the concept in question is wrong, and stands in need of improvement. However, what's important to note is that this does not involve an explicit intention to cause another to doubt their interpretive abilities.

I want to finish this section by raising a concern: Why would second order disagreement cause someone epistemic distress given that the dispute could be resolved by simply introducing a new term to express one of the concepts (e.g., 'sexual pestering' as opposed to 'sexual harassment')? In some cases, this might be the best strategy. It may be too difficult to convince someone to take up a new concept to be expressed by a word that has an established word-concept pairing. However, for a good deal of cases, we have an investment in words owing to their normative valence and function in our social practices – what Herman Cappelen (2018) calls 'lexical effects'. In particular, words like 'sexual harassment' do serious normative and explanatory work. People already have the term 'sexual harassment,' and it carries cognitive and affective associations (e.g., a serious moral wrong) that enable inferences to be made, informing our decisions-making practices (e.g., staying away from a perpetrator) and normative judgements (e.g., the perpetrator is the appropriate target of blame, resentment, legal action, etc.). And this investment in words can also partially explain why someone might be motivated to (intentionally) gaslight at the second order. A gaslighter might recognize the normative significance of the word under dispute, and, owing to this, will want to either keep the word or avoid it depending on his purposes.

One might wonder: Why bother distinguishing between first and second order gaslighting? First off, it is important to make as much sense of injustice as we can. When we think of gaslighting, it strikes me that we tend to see it as occurring only at the first order; so revealing the nature of second order gaslighting is useful insofar as we can identify its occurrence. But perhaps more importantly, first and second order gaslighting involve different wrongs, and require different solutions. The wrongs of first order gaslighting have been discussed in much detail in the philosophical literature (Abramson 2014; McKinnon 2017, forthcoming; Stark 2019). Thus, I will spend time explicating distinctive wrongs of second order gaslighting. I suggest that there are (at least) three: (1) *metalinguistic injustice,* (2) *conceptual obscuration,* (3) *perspectival subversion.* Each of which

[184] Second order gaslighting, which operates within unjust social environments, could be construed as a means of perpetrating *structural gaslighting* (Berenstain Forthcoming) – conceptual work that obscures the non-accidental connections between oppression and the harms they produce. In second order gaslighting, a gaslighter (typically) endorses an ideological concept that masks oppression, such as the ideological concept of sexual harassment endorsed by the man in the example we've been considering. Unfortunately, I don't have space to discuss this point further.

constitute a form of *epistemic* or *linguistic injustice.* I show how each of the injustices reliably have sequelae in terms of psychological and practical control.

## 7.6. Metalinguistic Injustice

As was discussed in the previous chapter, metalinguistic deprivation amounts to the following:

> **Metalinguistic deprivation** occurs when one is restricted, relative to their interlocutor, in their ability to contribute to the processes involved in determining the concept that will be expressed by a word in a context.

And metalinguistic deprivation can constitute metalinguistic injustice in three ways: When one is excluded from the processes involved in determining the *cognitive resource*s in a local context (EI); when one is excluded from the processes involved in determining the *linguistic* resources in a local context (LI); or when a speaker suffer from linguistic objectification owing to being metalinguistically deprived (LI*). How do each of these relate to gaslighting? Let me start with (EI) and (LI).

On the intentional story of gaslighting at the second order, a gaslighter might intend to induce doubt in the subject of gaslighting by saying things like, 'You women are hysterical,' 'You're probably on your rags,' 'You're so emotional,' and the like. In such cases, the gaslighter is attempting to make salient prejudicial stereotypes (i.e., cultural tools that narrow the range of expected behaviour) associated with a particular social category with the aim that the subject of gaslighting *come to believe that such stereotypes accurately represent who she is.* If the subject of gaslighting comes to form such a belief, this sets up the conditions for successful metalinguistic deprivation. The subject of gaslighting will feel that she cannot properly contribute to concept-determining conversation in virtue of doubting her interpretive abilities, which is a consequence of thinking that a pernicious stereotype aptly applies to herself. That is, the woman is unable to contribute to the epistemic or linguistic resources of a local context owing to the operation of identity-prejudice in that context – and this is due to the fact that the man intends for the woman to take such identity-prejudicial stereotypes to be apt or appropriate.

On the unintentional story, the gaslighter does not intend to induce doubt in the subject of gaslighting, but nonetheless brings to salience a prejudicial stereotype that causes the subject of gaslighting to doubt herself. This can be as simple as dismissing the testimony of women in contexts where they tend to be disbelieved (e.g., 'that's not sexual harassment' activates certain characterizations and classificatory dispositions). Again, the subject of gaslighting becomes restricted in her ability to contribute to concept-determining conversation in virtue of thinking that a pernicious stereotype aptly applies to herself.

Metalinguistic injustice is also a source of psychological and practical control. Broadly, I take 'control' to refer to the ability to influence someone contrary to their interests.[185] This can be innocent. I am interested in cases in which control, in this sense, is *unfair*. And I take it that when one is influenced contrary to their interests *in virtue of* suffering from epistemic injustice, that this constitutes unfair control.

We can say that *S* has *psychological* control over *T* when *S* is able to influence *T's* psychological states, such as how *T* feels and how *T* sees the world; and *S* has *practical* control over *T* when *S* is able

[185] This is close to the *power-over* concept of power. See Weber (1978).

to influence the way that *T* engages with their social environment. In cases of (second order) gaslighting, psychological and practical control is unfair. Let's discuss the former in more detail first.

Metalinguistic deprivation can be a source of psychological control insofar as the gaslighter can influence the beliefs that the subject of gaslighting has, and motivate certain affective attitudes. In the example that we've been exploring, the woman acquiesces to the concept of sexual harassment endorsed by her gaslighter. Because of this, she must come to accept certain things. She must accept John did not sexually harass her. Further, she must take on the attitude that she was wrong to accuse John of sexual harassment.[186] Accepting such things is apt for inducing certain affective states. The woman might feel anxious or insecure around her John, second-guess herself when brushed up against, feel unwelcome in certain spaces, and take herself to be 'bad' for ever thinking she was harassed, etc. The subject of gaslighting is under the psychological control of her gaslighter in virtue of being excluded from concept-determining conversation that causes her to take on attitudes, both cognitive and affective, that are contrary to her interest in being safe.

Metalinguistic injustice is also a source of practical control. This owes to the practical consequences of accepting a concept as apt or appropriate for a context. Consider the case earlier in the chapter where a wife catches her husband kissing another woman. Imagine that the husband denies that kissing counts as 'cheating'. Let's say that the husband suggests this because he recognizes that if he accepts his wife's preferred concept, he will be in the wrong. Now, suppose that the wife, owing to metalinguistic deprivation, accepts the husband's concept of cheating. A number of things practically follow from this. One consequence is that the wife must no longer interact with her husband as a cheater – she no longer has grounds for accusing him of wrongdoing, leaving him, asking for an apology, being emotionally distant, demanding that he change his ways, etc. And, because she may feel bad for accusing her husband of cheating, this might mean that she changes the way she relates to him, such as being overly apologetic, accommodating, and cautious not to anger or upset him. Moreover, the husband may continue to kiss other women, even in front of his wife, yet the wife feels that she isn't licensed to call him out. Other consequences abound. Through second order gaslighting, the gaslighter can push the subject of gaslighting to accept a concept, which determines the fittingness of certain actions for the context, thereby influencing the subject of gaslighting to relate to her social environment contrary to her interest in being respected by her partner. Gaslighting can motivate a target to perform actions that she otherwise wouldn't.

## 7.7. Conceptual Obscuration

A further wrong of second order gaslighting occurs when the subject of gaslighting forms the attitude that she has an inaccurate concept and, owing to this, *adopts the preferred concept of her gaslighter.* For example, the woman who is brushed up against may take on board her gaslighter's words that such an act is far too trivial to count as sexual harassment, and thus comes to accept that it does not constitute sexual harassment. Clearly, this has contextual implications. 'Sexual harassment' will express the gaslighter's preferred concept of *sexual harassment*. But it also has cross-contextual implications. The woman will become disposed to token her gaslighter's preferred concept in any context that she deems relevantly similar (i.e., any context in which she, or others, experience particular unwanted sexual advances). And given that the gaslighter's preferred concept of *sexual*

[186] And perhaps take on the attitude that she is unreliable when it comes to picking out sexual harassment; and that she is better off deferring to the judgment of her colleague. Though, this is probably more likely if she suffers from perspectival subversion (see §8.8).

*harassment* is inaccurate, this means that the woman will not recognize sexual harassment when it occurs. Because of this, the woman *loses important knowledge about the world*, and, thus, *is unable to communicate important information across social space.* Call this *conceptual obscuration.*

Conceptual obscuration undermines someone in their capacity as an epistemic/linguistic subject, and thus constitutes epistemic/linguistic injustice. How? The first thing to notice is that it involves someone *diminishing the knowledge of another by obscuring their epistemic/linguistic resource.* However, this does not yet constitute an epistemic/linguistic injustice – obscuring one's epistemic/linguistic resource might be quite innocent. For example, in a friendly conversation one might be told that couches are chairs, therefore obscuring their concept of *chair.* It is far from intuitive to say that this constitutes an injustice. To locate the injustice, we must look at *the means by which a cognitive resource is obscured.* And the means by which obscuring an epistemic/linguistic resource constitutes an injustice is through *metalinguistic deprivation.* It is one thing to be wronged by being denied full participation in concept-determining conversation, it is another thing to be wronged by being denied this *and* lose knowledge in the process (in virtue of adopting an inaccurate concept endorsed by an interlocutor). The wrong is located not simply in denying someone full participation in concept-determining conversation, but in the fact that this denial is causally responsible for someone losing knowledge about the world. Thus, we can define conceptual obscuration as:

> **Conceptual obscuration** is an injustice that occurs when *S* possesses an accurate concept *c,* but in virtue of suffering from metalinguistic injustice, *S* comes to replace *c* with a nearby inaccurate concept *c**.[187]

This is distinct from other forms of injustice. It is not hermeneutical injustice insofar as there is no lacuna in the collective hermeneutical resource. It is not *just* wilful hermeneutical ignorance or contributory injustice insofar as it isn't simply about imposing an ill-fitting concept onto marginalized experience. Instead, it is about one losing knowledge they once had, and the means by which this occurs.[188]

Losing knowledge entails that there are things to which the subject of gaslighting is no longer licensed. Like metalinguistic deprivation, this sets up the conditions for psychological and practical control. Consider again the example of the cheating husband. In this case, the wife accepts her husband's concept of cheating. When this occurs, she loses knowledge of what cheating is; she can no longer reliably pick it out. This means she is no longer licensed to believe her husband is cheating, or accuse him of doing so – her newly acquired concept doesn't take kissing to count. This constitutes psychological control insofar as the gaslighter is able to influence what the subject of gaslighting is able to *know* or *rationally believe.* Moreover, this translates into practical control. The wife's

[187] Conceptual obscuration can also occur when one possesses a *more* accurate concept, but in virtue of metalinguistic deprivation comes to adopt a *less* accurate concept. The idea is that the initial concept that is lost might not perfectly represent facts about the world, but it does a better job than the concept that is subsequently adopted.

[188] The primary function of many concepts isn't to reflect the world. As a result, they are not susceptible to standards of accuracy. Thus, for such concepts, conceptual obscuration is not a possibility. Thank you to an anonymous reviewer for this suggestion.

practical reasoning is affected insofar as she cannot, say, rationally choose to leave her husband, or confront him for his actions, given that he hasn't 'cheated' under the relevant definition.[189]

## 7.8. Perspectival Subversion

The final wrong of second order gaslighting that I will discuss is *perspectival subversion.* Again, this constitutes a form of injustice. It is not about being denied full participation in concept-determining conversation, or coming to adopt inaccurate concepts. Instead, it concerns the impairment of one's ability to *independently* interpret situations. In other words, it is a wrong that occurs when a subject of gaslighting is exposed to persistent challenges to her concepts over time, and because of this reaches a point where she forms the attitude (i.e., belief, anxiety) that her interpretive abilities are so unreliable that she must defer to her gaslighter's conceptual judgment (i.e., the right or appropriate concept to token). This comes in degrees. At the extreme end, it involves a total loss of self-trust. [190] Such cases are perhaps rare. More commonly, a subject of gaslighting will lose self-trust in particular domains – such as whether an event counts as sexual harassment. In any case, this captures a unique and pernicious wrong of gaslighting:

> **Perspectival subversion** is an epistemic injustice that occurs when a subject of gaslighting is the target of persistent conceptual challenges over time such that they come to doubt their ability to make conceptual judgments (in a domain, or across domains), and so they defer to the conceptual judgment of their gaslighter.[191]

Why 'perspectival'? It concerns how gaslighting affects one's *disposition* to construct interpretations of events (Camp 2018). Contrast this with the wrongs discussed above. Metalinguistic deprivation and conceptual obscuration involve the disposition to token a concept. In the former, it is the deployment of a concept that is challenged; in the latter, it is the deployment of a concept that is challenged, and a new concept is adopted to be used. Perspectival subversion, on the other hand, involves a disposition *not* to deploy concepts. It involves a subject of gaslighting refraining from making conceptual judgments because they take themselves to be unreliable (in a domain, or across domains) after being exposed to persistent challenges to their conceptual understanding. For example, consider the case above where the wife sees her husband kiss another woman. Let's say that in this marriage, the wife regularly sees her husband kiss other women, and when she confronts him about it each time, he consistently and repeatedly tells her that she doesn't really understand what 'cheating' is – he tells her that kissing doesn't count. He aggressively accuses her of being crazy when she brings it up, and says that she needs to stop being so sensitive. After a while of this treatment, the wife doubts the accuracy of her concept of cheating. However, she doesn't replace it with the one endorsed by her husband. Instead, she is so confused about what things count as cheating, and what things don't, that she decides her interpretive abilities aren't good enough to reliably pick it out, or to

[189] Another way to put this is that the woman would be in violation of (contested) knowledge norms if she accuses her husband of cheating. Because she doesn't 'know' that her husband cheated on her, she isn't licensed to: *assert* that he cheated on her (norm of assertion); use cheating in her *practical reasoning* (norm of action)*;* and form *beliefs* about her husband cheating (norm of belief). For more on knowledge norms, see: Williamson (2000) and Hawthorne and Stanley (2008).

[190] Including very mundane activities such as knowing how to catch a bus.

[191] How much one defers comes in degrees within domains.

discern the correct concept.[192] The wife becomes disposed not to make conceptual judgments about cheating.

Making a conceptual judgment is often necessary for practical reasons. For example, interpreting a situation as cheating allows one to act in response to the situation *as* cheating. Because of this, the subject of gaslighting must defer to the conceptual judgment of her gaslighter if she is to have some idea as to how to act 'appropriately' in a context. This is a form of psychological and practical control. It forces the subject of gaslighting to 'outsource' her interpretive abilities. Her judgment about an event is dependent on how her gaslighter sees things. Because the subject of gaslighting does not have sufficient control over her interpretation, owing to persistent challenges to her conceptual understanding, she is undermined in her capacity as an epistemic subject. She suffers from epistemic injustice.

## 7.9. Conclusion

The foregoing aimed to answer the following question: *In what sense does one doubt their understanding of reality when gaslit*? I argued that gaslighting can cause someone to doubt their understanding of reality in different ways depending on the linguistic order at which a gaslighting exchange takes place. This marks a distinction between *first order* and *second order* gaslighting. Given its omission in the philosophical literature, I focused on second order gaslighting: A form of gaslighting that involves inducing doubt in someone by targeting their conceptual understanding. I argued that it involves three wrongs: (1) *metalinguistic injustice,* (2) *conceptual obscuration,* and (3) *perspectival subversion.* Each reliably have sequelae in terms of psychological and practical control.

[192] More than this, the wife might be scared to face the wrath of her husband if she accuses him of cheating (again). This isn't directly connected to gaslighting, but poses a barrier to the wife's ability to build confidence in making conceptual judgements. Thus, her perspectival subversion becomes more difficult to escape.

# Chapter Eight

## Unjust Conversational Relationships

*What words do is not a simple function of either speaker intention or speaker awareness. [A] remark can oppress even in cases where the speaker has no special authority. Ordinary people under ordinary circumstances can unwittingly oppress others with their everyday comments. The power to verbal oppress comes from the social context; it need not reside in the speaker.*

Mary Kate McGowan 2019, p. 1

### 8. Introduction

The last three chapters all explored distinctive wrongs that occur within metalinguistic discourse. For example, metalinguistic injustice, however it is construed, concerns how a marginalized speaker suffers in an exchange over the concept that should be expressed by a shared word in a context, owing to identity prejudice. In this chapter, the final chapter, I will explicate one further distinctive wrong that occurs as a result of the operation of identity prejudice in metalinguistic discourse. However, this wrong goes *beyond* discussion about concepts. It infects the broader conversation. In other words, I will show how conceptual accommodation, either *Establishment* or *Negotiation*, can place speakers in an *unjust conversational relationship*.

### 8.1. Unjust Conversation

Mihaela Popa-Wyatt and Jeremy Wyatt (2017) offer a framework for understanding how slurring utterances can create unjust power imbalances through the assignment of discourse roles. They argue that since discourse is a social interaction, and social interactions are guided by agents conforming to certain social roles, then discourse roles are short-term social roles that exist for the purposes of governing conversational kinematics.

Discourse roles have three distinctive components: discourse status, discourse rules, and an associated social role. Discourse status concerns the relative rank between discourse roles within conversation; discourse rules regulate the expected and permissible moves of conversational participants, as well as providing rules for interpretation; and discourse roles have associated, long-term social roles that are inherited from broader social relations and made salient in conversation (2017, p. 2889). To get a better grip on this, consider an exchange:

A: Elms and beeches are the same kind of tree.
B: No, they're not!
A: How would *you* know?
B: I'm a dendrologist.
A: Ah, I guess you're right then.

During the course of this conversation we can see the discourse roles of A and B change. A's first utterance to B is assumed to be one made between peers; or, perhaps, more appropriately, both A and B occupy the discourse roles of peers. It is only until B reveals that she is an expert in wooded plants that discourse roles shift: A *defers* to B. After it is made known that B is an expert (discourse status), B is able to make certain moves in conversation relative to A (discourse rules), such as speaking authoritatively about wooded plants, in virtue of being assigned a particular role that is inherited from broader social relations and made salient in conversation (associated social role) – namely, being an expert in a domain.

To explain how slurring utterances create power imbalances via speech acts that assign discourse roles, Popa-Wyatt and Wyatt appeal to the work of Mary Kate McGowan. According to McGowan, various rules of accommodation make certain utterances count as correct play by enacting permissibility facts. McGowan calls such utterances 'conversational exercitives' (2009, 2012, 2018). Oppression, according to McGowan, is a rule-governed activity. It involves, *inter alia,* performing speech acts that determine permissible moves in a game of oppression. Hence, sexist and racist speech are exercitives that change what is permissible to say in conversation – they are moves that shape the character and direction of an oppression (language) game.

Extending McGowan's proposal, Popa-Wyatt and Wyatt suggest that we can add discourse roles as entities into score, and this contributes to the determination of correct play in conversation. When someone is the target of a slur, they are assigned, via an exercitive, a subordinate discourse role. And due to the accommodation of this exercitive, the score updates to include this assignment. This then changes discourse rules. It changes what is expected and permissible for the target of the assignment to say. Consider an example.

Suppose that Yu Yan and Eric are arguing over Australian immigration policy. Yu Yan, who is East Asian, believes that Australia should be more welcoming of foreigners; Eric, who is White, thinks that Australia should close its boarders. Let's imagine that Yu Yan says, 'My family came to this country decades ago, and they are hard working people,' to which Eric responds, 'Oh come on, you chinks are taking all of our jobs!' In this exchange, Yu Yan and Eric begin the conversation on, more or less, an even playing field – neither seem be assigned a discourse role that is hierarchically above the other.[193] However, Eric's use of a slurring utterance, which constitutes performance of an exercitive, assigns a new pair of discourse roles. Eric is assigned a powerful, dominant role; Yu Yan is assigned a subordinate role. That is, Eric has made it permissible to perform speech acts that are unavailable to Yu Yan. Eric can contribute to conversation as someone who belongs to the dominant White community, the members of which are entitled to certain jobs; Yu Yan must contribute to conversation as a foreigner or an 'invader,' who is a part of a community that denies entitlements to White Australians.

According to Popa-Wyatt and Wyatt, slurring utterances, in particular, are oppressive when one speaker gains unjust power over their interlocutor (Weber 1978; Dahl 1957). What is meant by power? '[P]ower is the extent to which a person A can affect a person B contrary to B's interests' (Popa-Wyatt and Wyatt 2017, p. 2880). And, it seems on Popa-Wyatt and Wyatt's account, acting contrary to someone's interest constitutes harm. Thus, 'A's power over B becomes oppressive if it is used to cause harm to B,' and the goal of a slurring utterance is 'to achieve and maintain unjust power

[193] This strikes me as controversial. I don't think that any marginalized person who enters into a conversation with a dominant speaker is on a level playing field. Certain roles are already operative, even if they aren't operating explicitly. This idea has been explored in chapter two, five, six, and seven.

over the target' (ibid. 2017, p. 2880). That is, slurs function 'as a mechanism of *unjust and forceful discourse power grab*' (ibid. 2017, p. 2895).

I accept that there are times at which an utterance constitutes a form of oppression when one speaker influences another contrary to their interests. However, I want to move away from understanding the injustice involved in conversation simply in terms of acting contrary to a speaker's interests. Instead, I will talk about how certain utterances establish conversational relations that unjustly constrains the *communicative agency* of a conversational participant – a form of agency to which speakers are entitled.

Communicative agency, according to Rachel Ann McKinney, is 'the ability of an agent to express attitudes to others... and the ability to do so *voluntarily*' (2016, p. 276). Denying one agency of this kind involves a distinctive wrong in which one is undermined in their status as a *communicative agent* – 'as people capable of choosing when to speak and what to say' (2016, p. 259). After all, according to McKinney, our agency is manifested through speaking. And the wrong of constraining communicative agency is characteristic of general practices of injustice, including manipulation, forced confession, coerced consent, etc. insofar as it 'rob[s] individuals... of the intentional agency normally required for communication' (ibid. 2016, p. 259).

Of course, it seems, at least intuitively, that there are cases of constrained speech that do not constitute injustice. For instance, a teacher might threaten a student with detention if the student doesn't reveal who started a schoolyard fight. Perhaps this is wrong, but it isn't clearly wrong. Thus, the kinds of cases of constrained speech that are clearly wrong will either involve the *means* by which speech is constrained, or its *consequences*. To see this, let's focus on a case offered by McKinney.

McKinney interest is in *extracted* speech. This occurs when a speaker is made to say something that they otherwise wouldn't – it involves cases in which one loses their ability to exercise agency with respect to the voluntariness of their speech. She uses the example of the forced confession to the rape and attempted murder of a 28 year old White woman of members of the 'Central Park Five.' The details of this case are grim. Young Black American and Latinx boys, ranging from 14 to 16 years old, were sent to jail on the grounds of confessions that were elicited via 'processes that significantly and wrongly undermined their agency... under conditions of structural violence' (ibid. 2016, p. 265). That is, the boys, who are subject to structural identity prejudice, were coerced and manipulated into confessing to a crime of which they were not guilty.

In the first place, we might appeal to the consequences of the forced confession to assess the wrong-making features of extracted speech. After all, the innocent boys went to jail. And further, this may contribute to perpetuating future discrimination against Black American and Latinx people in virtue of its contribution to stabilizing or regulating pernicious stereotypes (or generic characterizations). However, we can also appeal to the means by which such speech was extracted to assess its wrong-making features. As McKinney shows, the confession of the members of the Central Park Five was *forced* through coercion and manipulation. The communicative decision-making of each member was obscured by the immense power of legal institutions and racist social structures that weighed heavily against the innocent boys. Each of them was undermined in their capacity as communicative agents.

What we can see is that McKinney's account of power, which constrains communicative agency, is broadly in line with Miranda Fricker's (2007) notion of power *as* social control. Power is a capacity that operates either 'actively or passively, agentively or structurally, via control over agents' actions' (McKinney 2016, p. 265). Thus, this notion of power seems to deviate from the one offered by Popa-Wyatt and Wyatt insofar as it is possible that one can be subject to unjust power *even if* it is consistent

with their interests. After all, power relations can shape our desires and preferences.[194] Thus, the wrong I am interested in isn't so much to do with acting contrary to interests. Instead, it is about undermining a kind of agency to which people are entitled.

Extracting speech through coercion and manipulation is not the only unjust means of undermining the communicative agency of a speaker. I argue, in the spirit of Popa-Wyatt and Wyatt, that the communicative agency of a speaker can be undermined when she is in an *unjust conversational relationship*:

> S and T are in an **unjust conversational relationship** when S is assigned a subordinate discourse role, and hence is subject to unfair discourse rules, in virtue of T's conceptual move that applies a concept to S, which (i) brings to salience a social group that S belongs to, and (ii) invokes a subordinating stereotype (i.e., generic characterization) associated with that social group.

In other words, an unjust conversational relationship can be established in the following way. Through the performance of a conceptual move, T applies a particular concept to S which subordinates S in conversation. This is because the conceptual move made by T constitutes an exercitive that assigns to a subordinate discourse role to S. The discourse rules, which constitute the assigned subordinate discourse role, govern the moves S is expected and permitted to make, as well as how S's moves are to be interpreted.[195] In other words: S's communicative agency is constrained, and therefore undermined, by the assignment of a subordinate discourse role.

Moreover, the assignment of a subordinate discourse role, in order to constitute an *unjust* conversational relationship, must subject a speaker to *unfair* discourse rules. And, what makes discourse rules unfair, I argue, is that the discourse rules constrain the communicative agency of a speaker insofar as they are subject to 'targeted, persistent, and systematic control...' (McKinney 2016, p. 281) with respect to what they can contribute to conversation *in virtue of being a member of a social group that suffers from structural marginalization.*[196]

To see how all of this comes together, recall the role of *social environments* in the creation of oppressive contexts discussed in chapter five, §5.7. Social environments consist of background patterns that comprise a network of social relations. Such relations depend on the culturally available information that agents are able to draw on that frames expectation and enables (quasi) rule-governed interaction. That is, to coordinate at a given time, social agents draw on concepts and characterizations (i.e., stereotypes) to narrow the range of action-options, in a hierarchy of possibilities, to choose from. In unjust social environments, the patterns of interaction, undergirded by coordinated concepts, unfairly privilege some and subordinate others. Consider an example. Imagine the following exchange:

| | |
|---|---|
| Tanner: | Where are you from? |
| Flornita: | Melbourne. |
| Tanner: | No, where are you *really* from? |
| Flornita: | I'm Filipina. |

[194] For more on this idea, see Serene J. Khader (2011) on adaptive preferences.

[195] Perhaps also S is placed in a position of subordination in conversation whereby the determination of the speech act they perform is constitutively dependent (in part) by T's uptake. See Kukla (2014).

[196] One can be subject to unfair discourse rules for more agent specific reasons, such as when an older sibling has a greater conversational standing than a younger sibling. However, such cases aren't the interesting class of cases here.

Tanner: Wow, don't you guys eat dogs?!

Suppose this conversation takes place in a predominantly White community, that is, an unjust social environment constituted by a network of relations undergirded by social concepts that tend to operate within dominant White communities and that contribute to the marginalization of People of Colour. Tanner, whom we will imagine is White, intends to make salient a social group that Flornita belongs to (Asian/Filipino) in order to frame expectation about how to proceed in conversation. Initially, Flornita responds in a way that she wants the conversation to run. She intends to make salient the fact that she is from Melbourne. However, when Tanner asks where Flornita is *really* from, he makes a conceptual move, albeit covertly, that applies a particular concept to Flornita – perhaps something as general as the concept of foreigner, or something still general but a little more specific like the concept of Asian. That is, Tanner is performing a conceptual move that consists of classifying Flornita with a particular term (i.e., 'Asian') in order to shape the direction of conversation. Flornita is pressured to state that she is Filipina.

At this point, there is no assignment of discourse roles that places Tanner hierarchically above Flornita in conversation. However, in response to Flornita's claim that she is Filipina, Tanner makes another conceptual move. He applies a concept of Filipinx to Flornita that renders a pernicious stereotype salient in the context – namely, that in virtue of being Filipina, Flornita belongs to a community that harms and consumes dogs. Because of this, Flornita becomes a target of moral evaluation (condemnation) for Tanner. At this point, discourse roles change. Flornita is assigned a subordinate discourse role that is inherited from broader oppressive social relations and made salient in conversation: *Flornita's communicative agency is limited to a stereotype of what Tanner thinks Filipinx people are like*. Of course, this limitation can be resisted. However, resistance is often costly for marginalized folk, especially immigrants who wish to 'integrate' into dominant society. After all, the conversation between Flornita and Tanner is taking place in an unjust social environment. Thus, in the absence of resistance, Flornita is subject to control in conversation in virtue of belonging to a systematically marginalized social group; her communicative agency is undermined owing to identity prejudice.[197] In other words, Tanner and Flornita are in an unjust conversational relationship.

## 8.2. Conceptual Accommodation and Unjust Conversation: Establishment

What I will now discuss is how the accommodation of conceptual attitudes, in the context of real-word power structures, is often implicated in the creation of unjust conversational relationships.

To reiterate, *Establishment* (see chapter four, §4.6) is a form of conceptual accommodation in which a speaker makes a conceptual move, but the concept is not shared with the hearer. Thus, the hearer acquires the concept and takes on the attitude that the concept is apt or appropriate for the context.

In chapter four, I claimed that most cases of *Establishment* simply involve an exchange of information; and I discussed the example of intersectionality, where the conceptual interaction involved exchanging information about enquiry into gender relations. When a speaker says, 'Philosophical inquiry into gender relations should be intersectional,' and the hearer lacks possession of INTERSECTIONALITY, then the hearer will fail to fully grasp what was expressed by the utterance in

[197] Further, there is an asymmetry of power. Tanner does not belong to a systematically marginalized social group. Thus, Flornita does not have the means of constraining the communicative agency of Tanner in a way that is unjust and subordinating.

virtue of failing to understand how the word 'intersectionality' is being used. Thus, in order for the speaker's utterance to have its intended effect on score, the hearer must update their conceptual repertoire to include INTERSECTIONALITY, and take on the attitude that such a concept is apt or appropriate for the context.

In this case, very little seems to hinge on conceptual accommodation – in fact, it seems positive insofar as it improves an agent's conceptual repertoire. However, in some everyday conceptual interactions, those embedded in unjust power structures, introducing a new concept, and expecting a conceptual attitude to be shared, is not just a matter of exchanging information. It can be a mechanism of social control.

There are plenty of examples that we could bring to light, many of them newly coined concepts that feature in slurring utterances. For example, take the concepts FEMINAZI and SNOWFLAKE. Both are deployed by those who hold conservative political values, and used as a means of degrading progressives. The former is on the grounds that progressives are fascists. The latter is on the grounds that progressives are mentally unequipped to deal with criticism because they are too easily offended. In either case, I'll leave it to imagination as to how conversations might unfold when concepts of this kind are introduced. I want to move away from such examples to show that there are problematic cases of *Establishment* that don't involve concepts that feature in, or expressed by, typical slurring utterances.[198] Imagine the following conversation:

Connor: Where are you from?
Ligaya: I just moved here from the Philippines!
Connor: Oh, so you're a F.O.B!
Ligaya: What's that?
Connor: You know, immigrants who are 'fresh off the boat.'
Ligaya: I see. Well, I suppose I am…

Like the foregoing example, we can imagine that this exchange takes place in a predominately White community, let's say, an Australian context. Connor, who we will imagine is White, intends to make salient a social group that Ligaya belongs to (Asian/Filipinx) in order to determine how conversation should proceed. Ligaya obliges Connor's request for information and tells him that she is a recent immigrant from the Philippines. Upon hearing this, Connor makes a conceptual move: He introduces F.O.B. into conversation with the assumption that it will be understood. However, Ligaya does not possess this concept. Thus, Ligaya asks Connor for clarification; and Connor responds. When Ligaya comes to acquire the concept, she also takes on the attitude that such a concept is apt or appropriate for the context. Specifically, she accepts that the concept applies to her. This captures a case of *Establishment.* But, we might ask, what makes it problematic?

Given that the exchange takes place in a racist social environment (Australia), the introduction of the concept F.O.B., and the subsequent acceptance of its aptness or appropriateness, creates a pernicious insider/outsider distinction in the context. Connor belongs to the dominant, 'native'[199] White community; Ligaya is a member of the 'non-native,' immigrant community. We can imagine that this distinction inherits social meanings from broader practices, of which Ligaya is well aware. Immigrants are associated with invading the country, stealing jobs, being lazy, unhygienic, rude, and

[198] And therefore moving beyond the framework offered by Popa-Wyatt and Wyatt (2017).

[199] Native is in scare quotes because White people are not native to Australia – though, it is often believed that they are.

the like. Thus, while Ligaya may not intend for such cognitive and affective associations to enter into conversation, they nevertheless come along for the ride when F.O.B. is introduced. This is because, as discussed in chapter four, the acquisition of a concept often entails particular psychological effects. First, the characterization expressed by Connor produces *complicity* in Ligaya insofar as simply understanding the concept of F.O.B. requires that Ligaya structure her overall thinking in the same way as Connor. Second, taking on this frame of mind is *irresistible* insofar as a Ligaya will, more or less automatically, interpret a situation to match Connor's thoughts. And third, Ligaya cannot straightforwardly *deny* the concept of F.O.B. when it has been offered in conversation (Camp 2017, 2018).[200]

Thus, by accommodating Connor's conceptual attitude, including its psychological associations, Ligaya *establishes* the moves, conceptual and others, that can be made in conversation. Specifically, she renders Connor's conceptual move an exercitive that introduces a pair of discourse roles. Connor takes on a powerful role (i.e., dominant 'native'); and Ligaya takes on a subordinate role (i.e., 'invading' immigrant). This changes what is permissible for each of them to say, as well as how their utterances will be interpreted. Ligaya, who is assigned a subordinate discourse role, is unable to contribute to conversation in the way she prefers in virtue of being bound by the stereotype that has been introduced with a conceptual move (i.e., the use of F.O.B.).[201] *Establishment* puts Ligaya in a situation where she cannot help but be subjected to unfair discourse rules. This constitutes an unjust conversational relationship: Ligaya is assigned a subordinate discourse role, and hence is subject to unfair discourse rules, in virtue of Connor bringing to salience that Ligaya is an immigrant/Person of Colour, and invoking an associated negative stereotype. Until discourse roles change, Ligaya can only make moves in conversation as someone who falls under the concept F.O.B. *and* as someone to whom certain cognitive and affective (negative) associations are fitting (e.g., lazy, unhygienic, invader, etc.). The conversational relationship is unfair insofar as Ligaya is subject to discourse rules that constrain her communicative agency *in virtue of being an immigrant/Person of Colour*.[202]

## 8.3. Conceptual Accommodation and Unjust Conversation: Negotiation

In cases of *Negotiation*, conceptual accommodation occurs when speakers disagree as to which concept is apt or appropriate for a context, and one speaker comes to accept the concept endorsed by the other. I have argued that one speaker can have more influence in this process. And, when metalinguistic negotiation takes place within real-world power structures, this influence or control can be unjust. This is called *metalinguistic power*. When metalinguistic power operates within cases

[200] In addition to this, Rae Langton (2018) discusses a related notion that she calls *back-door-speech acts*: the use of a speech act that carries presuppositions that tend to enter into conversation by default. However, what I am suggesting isn't strictly to do with speech acts, or propositional presuppositions. Instead, it is about the introduction of concepts into conversation. This might come close to Christopher Potts' (2005) notion of *not-at-issue content*, however I won't discuss the similarities and differences here.

[201] Of course, it may be possible for the hearer to block the conceptual move by saying, 'This is not my word.' However, I am considering a case in which this block does not occur. The hearer might not decide to block for many different reasons – not wanting to cause a scene, not wanting to jeopardize their chance of 'integrating,' etc.

[202] It's important to note that the oppressive nature of such an exchange goes beyond facts about the conversational context. For instance, the use of F.O.B. might contribute to reinforcing unjust treatment against immigrants in the broader social environment.

of *Negotiation*, we might call this *Situated Negotiation.*[203] In chapter seven, I explored, in close detail, an example of this – a case where a man gaslights a woman into thinking that sexual harassment did not occur. Consider an extended version of the exchange:

| | |
|---|---|
| Woman: | You brushed up against my butt; that's sexual harassment! |
| Man: | Are you joking? Don't be so sensitive. |
| Woman: | Of course it is. It's unwanted sexual behavior. |
| Man: | You women are so sensitive.[204] |

To reiterate, this disagreement centers on the word 'sexual harassment' and the concept that it should express. And while we might be tempted to interpret the exchange as one concerning whether some shared criteria of sexual harassment has been met, this strikes me as wrongheaded. The woman and man seem to be disagreeing over the criteria that should be used to pick out or mark cases of sexual harassment. The man thinks that intentionally brushing up against someone is far too trivial. The woman resists this. There is a conceptual tug-of-war. I have argued that the woman might be disposed to accommodate the man's conceptual attitude partly in virtue of being situated in a patriarchal, sexist, and misogynistic social environment that fails to take seriously the testimony of sexual harassment and sexual assault survivors. The man has metalinguistic power, coeval with the woman's metalinguistic deprivation, in virtue of their respective social identities.

This is partly supported by empirical research in *Necessary Convergence Communication Theory* (Miller-Day 2004; Miller-Day and Jackson 2012; Dunbar and Mejia 2013). Findings in this area of research suggest that 'low-power partners in a relationship engage in "convergence communication" which is a pattern of uncritical acceptance or conformity to the more powerful partner's point of view' (Dunbar 2015, p. 8). It is thought that low-power partners aim to accommodate the more powerful partner's point of view in order to maintain the relationship, and this is linked to learned behaviors exhibited by participants of different genders shaped by common social assumptions and expectations (e.g., sex-roles).

With respect to the exchange above, we should expect that because gender is salient in the context, it will play a role in determining who has greater metalinguistic power. After all, the exchange takes place in an unjust social environment. Women are systematically subordinated, men are systematically privileged. Importantly, one form of socialization that we must bring attention to concerns women's historical association with hysteria. And, importantly, when this stereotype is raised, it can have a serious impact on a woman's self-conception. Ruth Bankey (2001) calls this *the fear of the hysterical woman*:

> The fear of the hysterical woman is a fear of being perceived by others as excessively feminine, out of control, and slipping into madness; therefore it is also a fear that one's experiences will not be taken seriously and that one is 'acting up' (2001, p. 37).

Bankey is bringing to light the idea that historical facts about women's association with hysteria and irrationality still contribute to their subordination today. In particular, it comes in the form of a fear; a fear of being pathologized and dismissed as 'out of their mind.' Thus, when the man brings to salience

[203] *Situated negotiation* is not a competitor to Plunkett and Sundell's (2013) notion of metalinguistic negotiation, but instead an extension of it.

[204] This is an instance of *gaslighting*, however, not all forms of negotiation are instances of gaslighting.

that his interlocutor is a woman, he is able to activate a pernicious, socially and historically embedded, stereotype that enacts permissibility facts in the conversation. That is, discourse roles are assigned: *he* is to be a trusted source of information; *she* is to be met with epistemic caution (Fricker 2007). The result is a power-base disequilibrium. An 'inequality of an individual's participation in and contribution to [understanding] in the interaction' (Dunbar 2015, p. 8).

From this, we can see how the exchange might unfold into an unjust conversational relationship:

Woman: I did overreact. I'm sorry.
Man: That's alright. Just don't be so uptight.

Given that the woman accepts the man's concept of sexual harassment is apt or appropriate for the context, further discourse roles are assigned. The woman's words can no longer be treated as reliable when it comes to whether the man, in fact, sexually harassed her. She loses her ability to say what she wants given that she is bound by the conceptual move introduced by the man that was made permissible partly in virtue of suffering from metalinguistic deprivation. The woman can only contribute to conversation *as* someone who was not sexually harassed, and as someone to whom cognitive and affective associations are fitting (e.g., epistemically unreliable). The conversational relationship, beyond metalinguistic discourse, is unfair insofar as the woman is subject to discourse rules that constrain her communicative agency *in virtue of being a woman.*

## 8.4. Comparisons

One might wonder: How is the wrong of an unjust conversational relationship different from the injustices discussed in the previous chapters? In chapter six, I discussed three ways of thinking about metalinguistic injustice: metalinguistic injustice (EI), metalinguistic injustice (LI), and metalinguistic injustice (LI*); and in chapter seven, I introduce two further injustices: conceptual obscuration and perspectival subversion. Surely the wrong of an unjust conversational relationship can be spelled out in terms of, at least, one of these? I will spend some time elaborating on their similarities and differences.

Metalinguistic injustice (EI) and (LI) involves preventing someone from contributing to the epistemic or linguistic resources of a local context. That is, one is restricted, either partially or wholly, from the processes that determine which concepts will be expressed by shared words in conversation. The wrong of an unjust conversational relationship is not about exclusion or denial, strictly speaking. Instead, the wrong of an unjust conversational relationship comes after this. It is about the unfair or illegitimate restriction of one's communicative agency, where one is assigned a subordinate discourse role, and is therefore subject to unfair discourse rules. Consider the foregoing example. In this case, the woman suffers from metalinguistic injustice (EI) and (LI); she is unable to contribute equally to *metalinguistic negotiation* owing to the operation of identity prejudice in the context. After this, however, the woman will still find herself in an on-going unjust conversational relationship. This is because the concepts that she takes as applying to herself, namely WOMAN and its packaging with HYSTERICAL, places the woman in a subordinate discourse role, and subjects her to unfair discourse rules that apply to her throughout the rest of the conversation (unless the rules change). Thus, the creation of an unjust conversational relationship is not just restricted to metalinguistic disagreements, but affects the woman's contributions to conversation more broadly. She may find herself unable to express doubt in her boss's testimony, accuse him of sexual harassment in the

future, or simply talk as epistemic peers in first order conversation. Of course, the creation of an unjust conversational relationship might assign unfair discourse rules that affect one's ability to contribute to *future* metalinguistic negotiation. However, the point I want to drive home is that the wrong of an unjust conversational relationship extends beyond the metalinguistic realm, unlike metalinguistic injustice (EI) and (LI).

Metalinguistic Injustice (LI*) involves denying someone the ability to exercise metalinguistic capacities. It occurs when a speaker suffers from linguistic objectification as a result of being metalinguistically deprived owing to the operation of identity-prejudice in a context. Taking lesson from what has just been said, one quick comparison we can make is that metalinguistic injustice (LI*) strictly concerns one's ability to contribute to metalinguistic disagreement, where the wrong of an unjust conversational relationship goes beyond this – it is about one being assigned a subordinate discourse role, and thus subject to unfair discourse rules, that affects their ability to contribute to conversation, generally speaking. However, metalinguistic injustice (LI*) is close to the wrong of an unjust conversational relationship insofar as it concerns *agency*. For metalinguistic injustice (LI*), the restriction of agency concerns undermining the ability of a speaker to perform certain metalinguistic moves. For unjust conversational relationships, the restriction is similar; it concerns what one can do in conversation once there is an ideological concept, or set of concepts, operative in the context. In this sense, the wrong of an unjust conversational relationship is a *generalization* of the wrong of metalinguistic injustice (LI*). We might say that the notion of an unjust conversational relationship is one way of spelling out linguistic injustice, broadly conceived.

Finally, the wrong of an unjust conversational relationship is distinct from conceptual obscuration and perspectival subversion. Conceptual obscuration occurs when a speaker possesses an accurate concept, but in virtue of suffering from metalinguistic injustice, the speaker comes to replace the concept with a nearby inaccurate concept. And perspectival subversion occurs when someone is the target of persistent conceptual challenges over time such that they come to doubt their ability to make conceptual judgments. What distinguishes these wrongs from an unjust conversational relationship is that neither involves assigning unfair discourse roles or rules. However, someone might suffer from conceptual obscuration or perspectival subversion *in virtue of* being in an unjust conversational relationship, or a series of unjust conversational relationships.

## 8.5. Conclusion

In this chapter, I explore one final wrong that occurs within metalinguistic discourse; and yet another social force that stifles conceptual innovation within conversation. This occurs when a speaker, in virtue of the identity category to which they belong, are placed in a subordinate position within conversation owing to the deployment of certain concepts and characterizations (i.e., stereotypes). I have called this an *unjust conversational relationship.* In particular, I discuss how unjust conversational relationships can be enacted within cases of conceptual accommodation. In some cases, an unjust conversational relationship can occur as a result of the acquisition of a concept, and subsequent construction of a subordinating characterization; and in other cases an unjust conversation relationship can occur as a result of a speaker 'losing' metalinguistic negotiation, or by coming to accept that a concept endorsed by an interlocutor is apt or appropriate for the context.

# Conclusion

The central task of this thesis was to offer support for the idea that we should reorient our understanding of the implementation challenge of conceptual engineering, at least for social kind concepts. In particular, we should be careful not to assume that individualistic implementation strategies, or those strategies that focus on the role of individual conceptual advocacy, will be the most appropriate means of bringing about conceptual change within a community. This is because there are many conservative social forces and infrastructure that work to maintain the conceptual status quo, and which stifle conceptual innovation. And such forces infect all aspects of our social lives, ranging from our formal institutions to everyday conversation and engagement with material reality. Thus, individuals will often be powerless, or severely constrained, in the face of structures that keep things conceptually as they are. There are many barriers to change that we must overcome.

Less abstractly, I identified two mechanisms of conceptual maintenance: *psychological convergence mechanisms* and *mechanisms of conceptual reproduction*. The former refers to the social practices that regulate similarity of characterization within a community; overlapping bodies of psychological information structured along the dimensions of prominence and centrality (i.e., stereotypes). Importantly, characterizations underlie higher order cognitive processes such as classificatory dispositions. Thus, changes to overlapping characterizations are apt to bring about changes to community-wide patterns of classification. And by revising community-wide patterns of classification, this is apt to bring about an ameliorative impact on social facts: we can engage in better linguistic and behavioural practices that correct unjust social relations (e.g., such as classificatory practices with 'woman' or 'man' that are trans* inclusive).

The latter mechanism of conceptual maintenance, which was the core focus of this thesis, refers to the social practices that directly stabilize community-wide classificatory dispositions. This is through social infrastructure that bolsters dispositions to copy prior stable patterns of usage with a term. Given that extant mechanisms of conceptual reproduction favour existing terms of conceptual engagement, we, as individuals, will find it difficult, if not implausible, to introduce new and better conceptual practices that will take root, propagate, and reproduce in dominant contexts.

After general discussion of these mechanisms, I mounted an argument against a particular individualistic implementation strategy that has gained some popularity in recent literature. This is the strategy of engaging in conversation, or more specifically disagreement, with respect to which standard of usage for a term should be operative in a context; or which concept should be expressed by a shared term in a context. This is called *metalinguistic disagreement*. I argued that there are strong social forces that govern the kinematics of such forms of disagreement that work to the advantage of existing patterns of classification. Put differently, speakers often have more control in the joint activity of pairing concepts with words in virtue of advocating for an extant and dominant conceptual practice. Importantly, this suppresses conceptual progress insofar as those who endorse alternative conceptual practices will often not be able to have their concept accepted as apt or appropriate for a context.

Apart from social forces rendering individualistic implementation strategies often implausible as a means of bringing about conceptual change, such forces are also typically caught up in identity related power relations. This means that individuals, in their attempt to implement a conceptual practice, are often subject to conditions of injustice. In particular, I argue that when identity prejudice plays role in determining who has greater control in the joint activity of pairing concepts

with words, this constitutes *metalinguistic injustice.* And there are different ways of thinking about the wrong-making features of metalinguistic injustice – as an epistemic injustice, a linguistic injustice, or both. In addition to this, metalinguistic injustice helps us to understand a unique form of gaslighting that I have called *second order gaslighting.* Finally, I argued that identity prejudice does not just affect the kinematics of metalinguistic discourse, but infects the character and direction of conversation more broadly. That is, the accommodation of conceptual attitudes, given the operation of identity prejudice in the context, can place speakers in an *unjust conversational relationship.* Thus, in order to successfully get another to adopt an engineered social kind concept in interpersonal speech situations, we must overcome not just the structural barriers that advantage dominant patterns of classification, but also structures of oppression.

With all of this in mind, one might wonder: which implementation strategies *should* we have? If social forces get in the way of conceptual innovation, how can we overcome such forces and bring about conceptual change? Unfortunately, I do not have time to explore possible solutions to the implementation challenge in great detail in this thesis. Part of the reason why is because this is a thoroughgoing empirical question: whether a particular implementation strategy will be successful depends on contingent facts about one's social environment (e.g., epistemic standing, social norms, structures of deference, pressure to conform, etc.). What I will do is point to historical examples that will give us clues as to which strategies for implementation might be likely to work.

One clear direction for developing better strategies for implementation is to look at the power of communities of conceptual practice. And such communities include consciousness-raising groups that serve as incubators for ideas within small in-groups, consisting of those who share, report, and aim to make sense of similar kinds of experiences. For example, Susan Brownmiller has discussed the historical development of the term 'sexual harassment':

> The 'this' they were going to break the silence about had no name. 'Eight of us were sitting in an office of Human Affairs,' Sauvigne remembers, 'brainstorming about what we were going to write on the posters for our speak-out. We were referring to it as 'sexual intimidation,' 'sexual coercion,' 'sexual exploitation on the job.' None of those names seems quite right. We wanted something that embraced a whole range of subtle and unstable persistent behaviors. Somebody came up with 'harassment.' Sexual harassment! Instantly we agreed. That's what it was (Brownmiller, Quoted in Fricker 2007, p. 149 – 150).

What we can see in this case is that there was missing conceptual practice, or what Miranda Fricker calls a 'hermeneutical lacuna' (2007, Ch. 7), that meant that women, in particular, were not able to make sense of their experience of sexual harassment, and to convey this experience across social space. The terms that were used prior to 'sexual harassment' did not fit properly with the women's inchoate thinking (i.e., characterization) about what had happened to them. And it was only until they entered into a particular community, exchanging experiences and ideas with others, that the term, and its psychological associations, was incubated, which allowed for the new conceptual practice to flourishing and propagate within feminist circles. And from here, the conceptual practice reproduced into dominant situatedness. It helped in legal challenges against perpetrators of sexual harassment, as well as helped shape laws to protect survivors.

Apart from conceptual practices being developed within communities, such practices can be antecedently developed, and then introduced into communities to be incubated, nurtured, and to be used more consistently. That is, a conceptual practice can enter into 'safe spaces,' or spaces that are

designed to be refuges from dominant life, as well as representations and manifestations of liberation and hope. For example, Patricia-Hill Collins talks about the role of safe spaces in the development of Black women's thought:

> These sites offered safe spaces that nurtured the everyday and the specialized thought of African-American women. In them Black women intellectuals could construct ideas and experiences that infused daily life with new meaning. These new meanings offered African-American women potentially powerful tools to resist the controlling images of Black womanhood. Far from being a secondary concern in bring about social change, challenging controlling images and replacing them with a Black woman's standpoint constituted an essential component in resisting intersection oppressions (Collins 2009, p. 123).

Collins shows us that safe spaces allow for marginalized thoughts to prosper; to develop into tools for fighting against structures of injustice.[205]

As alluded to in the Collins' quote, a conceptual practice might be developed within *academic* circles prior to entering into activist or non-academic communities. Again, this allows for the concept to flourish through sustained usage, and for its bugs to be fixed through exposure to focused and highly specific critical evaluation. Through this process, a conceptual practice is able to reach the point where it can be introduced and compete with dominant terms of conceptual engagement.[206] This is evident in the case of the concept of SPECICISM, its prevalence in animal rights communities, and its now common usage in ordinary language. The term 'speciesism' was first introduced to the academic world by philosopher Richard Ryder in the 1970s. The goal of Ryder, who now thinks there are multiple usages of 'speciesism,' was to capture the idea of there is a distinctive kind of wrong that occurs when one defends discrimination and exploitation by appealing to the fact that the subject of discrimination and exploitation is of a different species (Ryder 1998, p. 320). The term was then picked up by Peter Singer, and used in his hugely popular book *Animal Liberation* (1975), where his intention was to argue that merely belonging to a particular species should not be relevant to moral status. Now, after circulating within the academy, practices with the term 'speciesism' is common within activist circles (e.g., *PETA*, *Farm Animal Rights Movement*), and even with mainstream speakers. The conceptual practice had time to develop within technical spheres before it was adopted in more practical forms of life.

Yet another community-type response is more dialectic. It involves different communities influencing one another in the development of a conceptual practice. For example, think of intersectional analysis. This has a long history in Black feminism, consisting of activists and scholars (and activist-scholars) alike. Before the term 'intersectionality' was coined, there were a number attempts to name a form of oppression that emerges as a result of being situated along multiple identity axes, including the phrase 'interlocking systems of oppression' by the Black feminist lesbian socialist organisation the Combahee River Collective. Moreover, the political roots of intersectionality, in particular in the 1980s, can also be seen in the many publications of antiracist feminist texts in which the term 'intersections' was being somewhat regularly used (Carastathis 2014, p. 306). The term became hugely popularised when it was used by legal scholar Kimberlé Crenshaw

[205] Matthew J. Cull (forthcoming) has discussed the idea that we should look to history for precedents of conceptual engineering, and they offer a more extended discussion of Patricia Hill Collin's notion of a safe space.

[206] There is the risk that the conceptual practice will become too technical for ordinary practice, and thus less useful for practical forms of life.

in her academic and public writings, especially 'Mapping the Margins' (1991), and whose own activism in anti-violence lead her to recognise the complex interaction between different forms of oppression (ibid. 2014, p. 306).

So, what's the take away from all of this? It seems that what can help an alternative conceptual practice, those that challenge dominant terms of conceptual engagement, to compete with conservative social forces is to gain a critical mass of usage. And this mass of usage can be achieved through the development and nurturing of alternative conceptual practices within smaller communities that eventually grow over time through propagation and reproduction. Put differently, it seems that a conceptual practice stands a greater chance of entering into dominant situatedness when it reaches a point where the practice has sustained usage, within non-dominant communities, by a large enough group of people that can then spread the concept far and wide.

With this said, however, there are still major roadblocks that prevent the conceptual practices developed within non-dominant communities from being taken up in dominant contexts. This is to say that there are further structural barriers, keeping the conceptual status quo in place, that makes the critical mass necessary to bring about global conceptual change hard to meet. Such barriers are the means by which we produce and disseminate cognitive resources, such as classificatory practices and characterizations. This concern has been raised by theorists working on epistemic injustice, especially those interested in hermeneutical marginalization (e.g. Fricker 2007, Ch. 5.). What is at issue is the idea that dominant groups have far greater control over the processes by which conceptual practices are developed and circulated within a community of thinkers and speakers. This puts marginalized groups at a serious disadvantage when it comes to getting people to adopt a conceptual practice that makes better sense of marginalized experience.

More than this, recent literature in the epistemology of ignorance draws attention to the fact that our structures are geared toward supporting and maintaining *active ignorance*. Generally speaking, someone is in the grips of active ignorance when they have a tendency for vicious epistemic behaviour – this includes epistemic arrogance (i.e., a subject presumes that to know all there is), laziness (i.e., a lack of effort to find out more about the world), and close-mindedness (i.e., a lack of openness to others' experiences) (Medina 2013, p. 39).[207] Despite active ignorance being framed as a failure on the part of an epistemic agent, it is (often) a causal product of individuals being embedded in unjust hierarchical structures (Santos 2020, p. 4). In this sense, active ignorance is a social achievement. Specifically, it involves patterns of social practices infected by forms of prejudice that result in individuals being far more likely to come into contact with dominant perspectives, experiences, and understandings rather than the perspectives, experiences, and understandings that are unique to marginalized situatedness. And this inequality contributes to sustaining and bolstering systems of marginalization and oppression. So while active ignorance is enacted by individuals through cognition, it systematically arises as a part of structural processes (Woomer 2019, p. 77; Martín 2019, p. 12).[208] Put differently, unknowing is a consequence of effort; a product of social systems set

[207] See also Marilyn Frye (1983).

[208] Annette Martín (2019) distinguishes between different ways of thinking about structural ignorance, or White ignorance specifically. She argues that it is common to think about such ignorance in terms of willful ignorance (i.e., vicious epistemic behaviour) or the distribution of faulty cognitive resources (ibid. 2019: 6 – 11). However, Martín argues that neither view is sufficient to capture all of the ways that structural ignorance takes shape. She offers a general version of structural ignorance, or what she calls *The Structuralist view*, that states two conditions: such ignorance (1) results as part of systematic social processes (that give rise to injustice), and (2) that ignorance is an 'active player' in such processes (ibid. 2019: 12). In this sense, *The*

up in a way that allows for ignorance to reproduce unjust social relations. One need only think about the strong resistance to the phrase 'Black Lives Matter': this involves the systematic and willful ignorance of the violence against Black Americans at the hands of the supposedly protective institutions.[209]

How is ignorance a problem for conceptual innovation? Systematic ignorance will be brought about as a result of social forces that shape our minds to be attentive only to certain aspects of reality, while hiding others. This will include, among other things, psychological convergence mechanisms and mechanisms of conceptual reproduction. For example, anti-queer religious sex education in schools constrain the ability of students, typically those who are not queer, to appreciate and understand that gay, bi, lesbian people suffer a great deal of discrimination and persecution. This prevents individual thinkers from recognising the need for alternative ways of understanding the world; ways of understanding that would promote justice.

Given this, it seems that in order to bring about conceptual change, one thing that we can do to help is to target the structures that maintain active ignorance and which advantage dominant groups. That is, we can make efforts to reshape the way our conceptual practices are produced and disseminated. Linking to the discussion above, one thing that we can do is restructure our socio-epistemic landscape in a way that allows us to access more easily the conceptual practices that are operative within non-dominant communities. If our dominant structures were geared toward an openness to alternative communities of conceptual practice, the critical mass needed for a non-dominant conceptual practice to make its way into dominant contexts would be much lower; making conceptual innovation much more plausible.

One might respond that it is not enough to target our structures, such as mechanisms of conceptual reproduction and psychological convergence mechanisms. We are in the grips of ideology, and the problem is not just that people are embedded in unjust social processes of conceptual production and dissemination, but that people have vicious epistemic characters. People are epistemically arrogant, lazy, and close-minded. Subsequently, reshaping our social infrastructure will not be sufficient: we need people who will be open to new ways of thinking and who aim to understand the perspectives and experiences of those who are differently situated. As Medina argues, 'structural active ignorance can be corrected only by developing epistemic virtues such as epistemic humility, curiosity/diligence, and open-mindedness' (ibid. 2013, p. 23).

I agree with Medina that overcoming active ignorance requires that people cultivate better epistemic habits. However, this is not sufficient by itself. Of course, we need people to be epistemically humble, curious, and open-minded. But we need people to have these attitudes to, and with, the right concepts. It is no good having virtuous epistemic habits with the wrong concepts. And to help guarantee that people, across a community, have access to the right concepts, we need to have a well-regulated socio-epistemic environment.

I have not said enough to provide any concrete strategies for conceptual implementation. Of course, what will work is an empirical question. All I have done is gesture towards some possible answers: we can develop and nurture non-dominant conceptual practices within smaller communities in order for such practices to reach a critical mass needed to penetrate dominant contexts; and we can revise aspects of our social structure such that the critical mass needed to introduce a non-dominant conceptual practice into dominant contexts is far more easily achievable.

---

*Structuralist View* accommodates the former versions of ignorance, while being general enough to account for others.

[209] See also Charles Mills (1997, 2007).

I want to finish by saying that there are many more pernicious mechanisms of conceptual maintenance that must be identified if we are to come up with plausible implementation strategies. This thesis only scratches the surface. If we are serious about changing oppressive structures by targeting the concepts responsible for undergirding unjust social practices, then we must recognise all forms of conceptual maintenance that stand in our way.

# Bibliography

Abramson, Kate. 2014. Turning Up the Lights on Gaslighting. *Philosophical Perspective* 28 (1): 1 – 30.

Adams, Carol. J. 1990. *The Sexual Politics of Meat: A Feminist-Vegetarian Critical Theory.* Bloomsbury Press.

Adams, N. P. 2019. Authority, Illocutionary Accommodation, and Social Accommodation. DOI: 10.1080/00048402.2019.1621913.

Alcoff, Linda. 2010. Epistemic Identities. *Episteme* 7 (2): 128 – 137.

Allot, N. and M. Textor. 2012. Lexical Pragmatic Adjustment and Ad Hoc Concepts. *International Review of Pragmatics* 4 (2): 185 – 208.

Andreasen, Robin O. 2004. The Cladistic Race Concept: A Defense. *Biology and Philosophy* 19 (3): 425 – 442.

Appiah, Kwame Anthony. 1995. The Uncompleted Argument: DuBois and the Illusion of Race in *Overcoming Racism and Sexism,* Linda A. Bell and David Blumenfeld (eds.), pp. 59 – 78. Lanham, MD: Rowman and Littlefield.

Appiah, Kwame Anthony. 1996. Race, Culture, Identity: Misunderstood Connections in *Color Consciousness: The Political Morality of Race,* K. Anthony Appiah and Amy Guttmann, pp. 30 – 105. Princeton, NJ: Princeton University Press.

Balkin, Jack. 1998. *Cultural Software.* New Haven: Yale University Press.

Ball, Derek. 2020. Revisionary Analysis with Meaning Change (Or, Could Women Be Analytically Oppressed?) in *Conceptual Engineering and Conceptual Ethics*, Alexis Burgess, Herman Cappelen, and David Plunkett (eds.), pp. 35 – 58. Oxford: Oxford University Press.

Bankey, Ruth. 2001. La Donna é Mobile: Constructing the Irrational Woman. *Gender, Place and Culture: A Journal of Feminist Geography* 8 (1).

Barnes, Elizabeth. 2016. *The Minority Body: A Theory of Disability.* Oxford: Oxford University Press.

Barnes, Elizabeth. 2020. Gender and Gender Terms. *Noûs* 54 (3): 704 – 730.

Barsalou, Laurence W. 1989. Intraconcept Similarity and its Implications for Interconcept Similarity in *Similarity and Analogical Reasoning*, S. Vosniadou and A. Ortony (eds.), pp. 76 – 121. Cambridge: Cambridge University Press.

Barasalou, Laurance W. 2003. Abstraction in Perceptual Symbol Systems. *Philosophy Transactions of the Royal Society of London: Biological Sciences* 358: 1177 – 1187.

Berenstain, Nora. 2020. White Feminist Gaslighting. *Hypatia* 35 (4): 722 – 758.

Bicchieri, Cristina and Peter McNally. 2018. Shrieking Sirens: Schemata, Scripts, and Social Norms. How Change Occurs. *Social Philosophy & Policy* 35 (1): 23 – 53.

Brownmiller, Susan. 1990. *In Our Time: Memoir of Revolution.* New York: Dial Press.

Burge, Tyler. 1979. Individualism and the Mental. *Midwest Studies in Philosophy* 4: 73 – 121.

Burgess, Alexis and David Plunkett. 2013. Conceptual ethics I. *Philosophy Compass* 8 (12): 1091–1101.

Burgess, Alexis and David Plunkett. 2020. On the Relationship Between Conceptual Engineering and Conceptual Ethics. *Ratio* 33 (4): pp. 281 – 294.

Butler, Judith. 1998. *Excitable Speech: A Politics of the Performative*, New York: Routledge.

Camp, Elisabeth. 2013. Slurring Perspectives. *Analytic Philosophy* 54 (3): 330 – 349.

Camp, Elisabeth. 2015. Logical Concepts and Associative Characterizations in *The Conceptual Mind*, Eric Margolis and Stephen Laurence (eds.), pp. 591 – 621. MIT Press.

Camp, Elisabeth. 2017a. Perspectives in Imaginative Engagement With Fiction. *Philosophical Perspectives* 31. DOI: 10.1111/phpe12102.

Camp, Elisabeth. 2017b. Why Metaphors Make Good Insults: Perspectives, Presupposition, and Pragmatics. *Philosophical Studies* 174 (1): 47 – 64.

Camp, Elisabeth. 2018. Perspectives in Imaginative Engagement with Fiction. *Philosophical Perspectives* 31 (1): 73 – 102.

Cantalamessa, Elisabeth Amber. 2019. Disability Studies, Conceptual Engineering, and Conceptual Activism, *Inquiry* DOI: 10.1080/0020174X.2019.1658630.

Cappelen, Hermen. 2018. *Fixing Language: An Essay on Conceptual Engineering.* Oxford: Oxford University Press.

Cappelen, Herman and David Plunkett. 2020. A Guided Tour of Conceptual Engineering and Conceptual Ethics in *Conceptual Engineering and Conceptual Ethics*, Alexis Burgess, Herman Cappelen, and David Plunkett (eds.), pp. 1- 26. Oxford: Oxford University Press.

Cappelen, Herman and Ernie Lepore. 2007. *Language Turned on Itself: The Semantics and Pragmatics of Metalinguistic Discourse.* Oxford: Oxford University Press.

Cappelen, Herman and Josh Dever. (Forthcoming). On the Uselessness of the Distinction Between Ideal and Non-Ideal Theory (at least in the Philosophy of Language) in *Routledge Companion to Social and Political Philosophy of Language*, Justin Khoo and Rachel Sterken. Abingdon: Routledge.

Carastathis, Anna. 2014. The Concept of Intersectionality in Feminist Theory. *Philosophy Compass* 9 (5): 304 – 314.

Carey, Susan. 1985. *Conceptual Change in Childhood*. Cambridge, MA: MIT Press.

Carey, Susan. 2009. *The Origins of Concepts*. Oxford: Oxford University Press.

Carston, R. 1997. Enrichment and Loosening: Complementary Process in Deriving the Proposition Expressed. *Linguistische Berichte* 8: 103 – 127.

Chalmers, David. Forthcoming. What is Conceptual Engineering and What Should it Be? *Inquiry* DOI: 10.1080/002016X.2020.1817141.

Clark, Andy and David Chalmers. 1998. The Extended Mind. *Analysis* 58 (1): 7 – 19.

Collins, Patricia-Hill. 2009. *Black Feminist Thought: Knowledge, Consciousness, and the Power of Empowerment.* New York: Routledge.

Collins, Stephanie. 2019. *Group Duties: Their Existence and Their Implications for Individuals.* Oxford: Oxford University Press.

Crenshaw, Kimberlé. 1991. Mapping the Margins: Intersectionality, Identity Politics, and Violence Against Women of Color. *Stanford Law Review* 43 (6): 1241 – 1299.

Crerar, Charlie. 2016. Taboo Hermeneutical Injustice, and Expressively Free Environments. *Episteme* 13 (2): 195 – 207.

Crocker, Jennifer, Susan T. Fiske, and Shelley E. Taylor. 1984. Schematic Bases of Belief Change in *Attitudinal Judgment,* J. Richard Eiser (ed.). New York: Springer.

Cudd, Ann. 2006. *Analyzing Oppression.* Oxford: Oxford University Press.

Cull, Matthew J. Forthcoming. Engineering is Not a Luxury: Black Feminists and Logical Positivists on Conceptual Engineering. *Inquiry* DOI: 10.1080/0020174.1883476

Cumming, Samuel. 2013. From Coordination to Content. *Philosophers' Imprint* 13.

Dahl, Robert A. 1957. The Concept of Power. *Behavioral Science* 2 (3): 201 – 215.

Davies, Alex. 2020. Identity Display: Another Motive for Metalinguistic Disagreement. *Inquiry*. DOI: 10.1080/0020174X.2020.1712229.

Dembroff, Robin. 2016. What is Sexual Orientation? *Philosophers' Imprint* 16.

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

Dembroff, Robin. Forthcoming. Real Talk on the Metaphysics of Gender. *Philosophical Topics.*

Dembroff, Robin and Daniel Wodak. 2018. He/She/They/Ze. *Ergo* 5.

Deutsch, Max. 2020. Speaker's Reference, Stipulation, and a Dilemma for Conceptual Engineers. *Philosophical Studies.* DOI: 10.1007/s11098-020-01416-z.

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

Dotson, K. 2011. Tracking Epistemic Violence, Tracking Practices of Silencing. *Hypatia* 26 (2): 236 – 257.

Dotson, Kristie. 2012. A Cautionary Tale: On Limiting Epistemic Oppression. *Frontiers: A Journal of Women Studies* 33 (1), 24 – 47.

Dunbar, Norah. E. 2015. A Review of Theoretical Approaches to Interpersonal Power. *The Review of Communication* 15 (1): 1 – 18.

Dunbar, Norah. E. and R. Mejia. 2013. A Qualitative Analysis of Power-Based Entertainment and Interactional Synchrony in Couples. *Personal Relationships* 20 (3): 391 – 405.

Fischer, Eugen. 2020. Conceptual Control: On the Feasibility of Conceptual Engineering. *Inquiry* DOI: 10.1080/0020174x.2020.1773309.

Foucault, Michel. 1980. *Power/Knowledge: Selected Interviews and Other Writings by Michel Foucault,* 1972 – 77. C. Gordon (ed.) Brighton, England: Harvester.

Frege, Gottlob. 1892. On Concept and Object in *Translations From the Philosophical Writings of Gottlob Frege*, P. Geach and M. Black (eds.), 3rd Edition, pp. 42 – 55. Oxford: Oxford University Press.

Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.

Frye, Marilyn. 1983. *The Politics of Reality*. CA: Crossing Press.

Gerken, Mikkel. 2019. Pragmatic Encroachment and the Challenge from Epistemic Injustice. *Philosophers' Imprint* 19.

Gooding-Williams, Robert. 1998. Race, Multiculturalism, and Democracy. *Constellations* 5 (1): 18 – 41.

Gopnik, A. and A. Meltzoff. 1997. *Words, Thoughts, and Theories*. Cambridge, MA: MIT Press.

Hacking, Ian. 1995. *Rewriting the Soul: Multiple Personality and the Sciences of Memory*. Princeton, NJ: Princeton University Press.

Hall, Stuart. 1996. The Problem of Sociology in *Stuart Hall: Critical Dialogues in Cultural Studies*, Kuan-Hsing Chen and David Morely (eds.), pp. 24 – 45. New York: Routledge.

Hampton, James A. 2006. Concepts as Prototypes in *Psychology of Learning and Motivation,* B.H. Ross (ed.), pp. 79 – 113. New York: Academic Press.

Hansen, Nat. 2019. Metalinguistic Proposals, *Inquiry* DOI: 10.1080/0020174X.2019.1658628.

Hartstock, Nancy. 1998. *The Feminist Standpoint Revisited and Other Essays.* Colorado: Westview Press.

Haslanger, Sally. 1995. Ontology and Social Construction. *Philosophical Topics* 23 (2): 95 – 125.

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

Haslanger, Sally. 2012. Resisting Reality: Social Construction and Social Critique. Oxford: Oxford University Press.

Haslanger, Sally. 2015. Social Structure, Narrative and Explanation. *Canadian Journal of Philosophy* 45 (1): 1 – 15.

Haslanger, Sally. 2016. What is a (Social) Structural Explanation? *Philosophical Studies* 173: 113 – 130.

Haslanger, Sally. 2018. What is a Social Practice? *Royal Institute of Philosophy* 82, Supplement: 231 – 247.

Haslanger, Sally. 2019. Cognition as a Social Skill. *Australasian Philosophical Review* 3 (1): 5 – 25.

Haslanger, Sally. 2020. Going On, Not the Same Way in *Conceptual Engineering and Conceptual Ethics*, Alexis Burgess, Herman Cappelen, and David Plunkett (eds.), pp. 230 – 260. Oxford: Oxford University Press.

Hawthorne, John and Jason Stanley. 2008. Knowledge and Action. *Journal of Philosophy* 105: 571 – 590.

Hesni, Samia. 2018. Illocutionary Frustration. *Mind* 127: 947 – 976.

Hill-Collins, Patricia. 2002. Contemporary Black Feminist Thought in *African-American Philosophy: Selected Readings*, Tommy Lee Lott (ed.). NJ: Prentice-Hall.

Hornsby, Jennifer and Rae Langton. 1998. Free Speech and Illocution, *Legal Theory* 4 (1): 21 – 37.

Ichikawa, Jonathan Jenkins. 2020. Contextual Injustice. *Kennedy Institute of Ethics* 30 (1): 1 – 30.

Isaac, Manuel Gustavo. 2020. How to Conceptual Engineer Conceptual Engineering. *Inquiry* DOI: 10.1080/0020174X.2020.1719881.

Jenkins, Katharine. 2016. Amelioration and Inclusion: Gender Identity and the Concept of Woman. *Ethics* 126: 396 – 421.

Jones, Karen. 2008. How to Change the Past in *Practical Identity and Narrative Agency*, Kim Atkins and Catriona Mackenzie (eds.). Routledge.

Jones, Karen. 2012. The Politics of Intellectual Self-Trust. *Social Epistemology* 26 (2): 237 – 251.

Jones, Karen, Francois Schroeter, and Laura Schroeter. 2020. Mind-Making, Affective Regulation, and Resistance. *Australian Philosophical Review* 3 (1): 86 – 89.

Jorem, Sigurd. Forthcoming. Conceptual Engineering and the Implementation Problem. *Inquiry*.

Kahan, Dan, Donald Braman, John Gastil, Paul Sloiv, and C.K. Mertz. 2007. Culture and Identity-Protective Cognition: Explaining the White Male Effect in Risk Perception. *Journal of Empirical Legal Studies* 4 (3): 465 – 505.

Keil, Frank C. 1989. *Concepts, Kinds, and Cognitive Development*. Cambridge, MA: MIT Press.

Khader, Serene. 2011. *Adapative Preferences and Women's Empowerment.* New York: Oxford University Press.

Kirsch, Julie. 2019. Interpreting Our Emotions. *Ratio* DOI: 10.1111/rati.12251.

Kitcher, Philip. 2007. Does 'Race' Have a Future? *Philosophy and Public Affairs* 35 (4): 293 – 317.

Koch, Steffen. Forthcoming a. The Externalist Challenge to Conceptual Engineering. *Synthese* DOI: 10.1007/s11229-018-02007-6.

Koch, Steffen. Forthcoming b. Engineering What? On Concepts in Conceptual Engineering. *Synthese* DOI: 10.1007/s11229-020-02868-w.

Komatsu, Lloyd. K. 1992. Recent Views of Conceptual Structure. *Psychological Bulletin* 112: 500 – 526.

Kripke, Saul. 1970/1980. *Naming and Necessity*. Oxford: Basil Blackwell Ltd.

Kukla, Quill. 2014. Performative Force, Convention, and Discursive Injustice. *Hypatia* 29 (2): 440 – 457

Kukla, Quill. 2018. Slurs, Interpellation, and Ideology. *Southern Journal of Philosophy* 56 (S1): 7 – 32.

Langton, Rae. 1993. Speech Acts and Unspeakable Acts. *Philosophy & Public Affairs* 22 (4): 293 – 330.

Langton, Rae. 1998. Subordination, Silence, and Pornography's Authority in *Censorship and Silencing: Practices of Cultural Regulation,* Robert C. Post (ed.), pp. 261 – 283. Los Angeles, CA: Getty Research Institute for the History of Art and the Humanities.

Langton, Rae. 2012. Beyond Belief: Pragmatics in Hate Speech and Pornography in *Speech and Harm: Controversies Over Free Speech*, M. K. McGowan and I. Maitra (eds.). Oxford: Oxford University Press.

Langton, Rae. 2014. Hate Speech and the Epistemology of Justice: Review of Jeremy Waldron, *The Hate in Hate Speech. Criminal Law and Philosophy* 10 (4): 865 – 673.

Langton, Rae. 2015. Accommodating injustice, Locke Lectures at Oxford University.

Langton, Rae. 2018. Blocking as Counter-Speech in *New Work on Speech Acts*, D. Harris, D. Fogal, and M. Moss (eds.). New York: Oxford University Press.

Langton, Rae and Caroline West. 1999. Scorekeeping in a Pornographic Language Game, *Australasian Journal of Philosophy* 77 (3): 303 – 319.

LeDoux, Joseph E. and Richard Brown. 2017. A Higher-Order Theory of Emotional Consciousness. *PNAS* DOI: 10.1073/pnas.1619316114.

Lewis, David. 1979. Scorekeeping in a Language Game. *Journal of Philosophical Logic* 8 (3): 339 – 359.

Lindauer, Matthew. 2020. Conceptual Engineering and Concept Preservation. *Ratio* 00: 1 – 8.

Machery, Edouard. 2009. Doing Without Concepts. Oxford: Oxford University Press.

MacKinnon, Catherine. 1987. *Feminism Unmodified: Discourses on Life and* Law. Cambridge, MA: Harvard University Press.

MacKinnon, Catherine. 1993. *Only Words.* Cambridge, MA: Harvard University Press.

Maitra, Ishani. 2012. Subordinating Speech 1 in *Speech and Harm: Controversies Over Free Speech*, I. Maitra and M. K. McGowan (eds.). Oxford: Oxford University Press.

Maitra, Ishani. 2017. Speech and Silencing in *The Routledge Companion to Feminist Philosophy*, Ann Garry, Serene J. Khader, and Alison Stone (eds.). New York: Routledge.

Mallon, Ron. 2016. *The Construction of Human Kinds.* Oxford: Oxford University Press.

Manne, Kate. 2018. *Down Girl: The Logic of Misogyny*. Oxford: Oxford University Press.

Margolis, Eric and Stephen Laurence. 1999. Concepts and Cognitive Science in *Concepts: Core Readings*, Eric Margolis and Stephen Lawrence (eds.). Cambridge, MA: MIT Press.

Marques, Teresa. 2020. Amelioration vs. Perversion in *Shifting Concepts: The Philosophy and Psychology of Conceptual Variability*, Teresa Marques and Asa Wikforss (eds.). Oxford: Oxford University Press

Martín, Annette. 2020. What is White Ignorance? *The Philosophical Quarterly*. DOI: 10.1093/pq/pqaa073.

McGlynn, Aiden. 2019. Epistemic Objectification as the Primary Harm of Testimonial Injustice. *Episteme*: 1 – 17. DOI: 10.1017/epi.2019.9

McGowan, Mary Kate. 2009. Oppressive Speech. *Australasian Association of Philosophy* 87 (3): 389 – 407.

McGowan, Mary Kate. 2012. On 'Whites Only' Signs and Racist Hate Speech: Verbal Acts of Racial Discrimination in *Speech and Harm: Controversies Over Free Speech,* Mary Kate McGowan and Ishani Maitra (eds.). Oxford: Oxford University Press.

McGowan, Mary Kate. 2017. On Multiple Types of Silencing in *Beyond Speech*, Mari Mikkola (ed.). Oxford: Oxford University Press.

McGowan, Mary Kate. 2018. On Cover Exercitives: Speech and the Social World in *New Work on Speech Acts,* Daniel Fogal, Daniel W. Harris, and Matt Moss (eds.). Oxford: Oxford University Press.

McGowan, Mary Kate. 2019. *Just Words: On Speech and Hidden Harm*. Oxford: Oxford University Press.

McKinney, Rachel. A. 2016. Extracted Speech. *Social Theory and Practice* 42 (2): 258 – 284.

McKinnon, Rachel. 2017. 'Allies' Behaving Badly: Gaslighting as Epistemic Injustice in *The Routledge Handbook of Epistemic Injustice*, Ian James Kidd, Gaile Pohlhaus Jr., and José Medina (eds.) London: Routledge.

McKinnon, Rachel. Forthcoming. Gaslighting as Epistemic Violence: 'Allies,' Mobbing, and Complex Posttraumatic Stress Disorder, Including a Case Study of Harassment of Transgender Women in Sport in *Overcoming Epistemic Injustice: Social and Psychological Perspectives,* Benjamin Sherman and Stacey Goguen (eds.), pp. 285 – 301. Rowman and Littlefield.

Medina, José. 2013. *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and Resistant Imaginations.* Oxford: Oxford University Press.

Meibauer, Jörg. 2019. What is an Indirect Speech Act? Reconsidering the Literal Force Hypothesis. *Pragmatics & Cognition* 26 (1): 61 – 84.

Mikkola, Mari. 2009. Gender Concepts and Intuitions. *Canadian Journal of Philosophy* 39 (4): 559 – 583.

Miller-Day, Michelle. 2004. *Communication Among Grandmothers, Mothers, and Adult Daughters: A Qualitative Study of Maternal Relationships.* Mahwah, NJ: Lawrence Erlbaum.

Miller-Day, Michelle and A. W. Jackson. 2012. The Convergence Communication Scale: Development and Evaluation of an Assessment of Interpersonal Submission. *Journal of Social and Personal Relationships* 29 (8): 1036 – 1057.

Mills, Charles. 1997. *The Racial Contract.* Ithaca, NY: Cornell University Press.

Mills, Charles. 1998. *Blackness Visible: Essays on Philosophy and Race.* Ithaca, NY: Cornell University Press.

Mills, Charles. 2007. White Ignorance in *Race and Epistemologies of Ignorance*, Shannon Sullivan and Nancy Tuana (eds.), pp. 11 – 38. New York: State University of New York Press.

Murphy, Gregory L., and Douglas L. Medin. 1985. The Role of Theories in Conceptual Coherence. *Psychological Review* 92: 289 – 316.

Nado, Jennifer. 2019. Conceptual Engineering, Truth, and Efficacy. *Synthese* DOI: 10.1007/s11229-019-02096-x.

Nado, Jennifer. 2020. Taking Control: Conceptual Engineering Without (Much) Metasemantics. *Inquiry* DOI: 10.1080/002174X.2020.1850342.

Nado, Jennifer. Manuscript. Classification Procedures as the Targets of Conceptual Engineering.

Nussbaum, Martha. 1995. Objectification. *Philosophy and Public Affairs* 24 (4): 249 – 291.

Outlaw, Lucius. 1990. Toward a Critical Theory of Race in *Anatomy of Racism*, D.T. Goldberg (ed.). Minneapolis: University of Minnesota Press.

Pinder, Mark. Forthcoming a. Conceptual Engineering, Metasemantic Externalism, and Speaker-Meaning. *Mind.*

Pinder, Mark. Forthcoming b. Conceptual Engineering, Speaking-Meaning, and Philosophy. *Inquiry* DOI: 10.1080/0020174X.2020.1853342.

Perry, John. 1980. A Problem About Continued Belief. *Pacific Philosophical Quarterly* 61 (4): 317.

Perry, John. 2001. *Reference and Reflexivity.* Palo Alto, CA: CSLI Publications.

Plunkett, David and Tim Sundell. 2013. Disagreement and the Semantics of Normative and Evaluative Terms. *Philosophers' Imprint,* 13.

Podosky, Paul-Mikhail Catapang. 2018. Hermeneutical Injustice and Animal Ethics: Can Nonhuman Animals Suffer From Hermeneutical Injustice? *Journal of Animal Ethics* 8 (2): 216 – 228.

Podosky, Paul-Mikhail Catapang. 2018. What Defines a Conceptual Resource? *Ergo* 6.

Podosky, Paul-Mikhail Catapang. 2018. Privileged Groups and Obligation: Engineering Oppressive Concepts. *Journal of Applied Philosophy* DOI: 10.1111/japp.123998.

Podosky, Paul-Mikhail Catapang. 2018. What Defines a Conceptual Resource? *Ergo* 6.

Pohlhaus Jr., Gaile. 2012. Relational Knowing and Epistemic Injustice: Toward a Theory of Willful Hermeneutical Ignorance. *Hypatia* 27 (4): 715 – 735.

Pohlhaus Jr. Gaile. 2014. Discerning the Primary Harm in Cases of Testimonial Injustice. *Social Epistemology* 28 (2): 99 – 114.

Pollock, Joey. 2020. Conceptual Engineering and Semantic Deference. *Studia Philosophical Estonia* 12: 81 – 98.

Popa-Wyatt, Mihaela. and Jeremy L. Wyatt. 2018. Slurs, Roles and Power. *Philosophical Studies* 175: 2879 – 2906.

Potts, Christopher. 2005. *The Logic of Conversational Implicatures*. Oxford: Oxford University Press.

Putnam, Hilary. 1975. The Meaning of 'Meaning' in *Philosophical Papers Vol. 2. Mind, Language, and Reality,* H. Putnam (ed.), pp. 215 – 271. Cambridge: Cambridge University Press.

Richard, Mark. 2020. The A-project and the B-project in *Conceptual Engineering and Conceptual Ethics*, Alexis Burgess, Herman Cappelen, and David Plunkett (eds.), pp. 359 – 378. Oxford: Oxford University Press.

Riggs, Jared. 2019. Conceptual Engineers Shouldn't Worry About Semantic Externalism. *Inquiry* DOI: 10.1080/0020174x.2019.1675534.

Rosch, Eleanor. 1978. Principles of Categorization in *Cognition and Categorization*, E. Rosch and B. Lloyd (eds.), pp. 27 – 48. Hillsdale, NJ: Lawrence Erlbaum Associates.

Rosch, Eleanor and Carolyn B. Mervis. 1975. Family Resemblances: Studies in the Internal Structure of Categories. *Cognitive Psychology* 7: 573 – 605.

Rothbart, Myron. 1981. Memory Processes and Social Beliefs in *Cognitive Processes in Stereotyping and Intergroup Behavior*. Hillsdale, NJ: Erlbaum.

Rumelhart, David E. and Donald A. Norman. 1978. Accretion, Tuning and Restructuring: Three Modes of Learning in *Schematic Factors in Cognition,* J.W. Cotton and R. L. Klatskey (eds.). Hillsdale, NJ: Erlbaum.

Ryder, Richard. 1998. Speciesism in *Encyclopedia of Animal Rights and Animal Welfare*, M. Bekoff and C. Meaney (eds.), pp. 320. Chicago: Fitzroy Dearborn.

Santos, Breno R. G. 2020. Echo Chambers, Ignorance and Domination. *Social Epistemology* DOI: https://doi.org/10/1080/02691728.2020.1839590

Saul, Jennifer. 2006. Gender and Race. *Aristotelian Society Supplementary Volume* 55 (4): 119 – 143.

Saul, Jennifer. 2007. Philosophical Analysis and Social Kinds: Gender and Race. *Proceedings of the Aristotelian Society* 106 (1): 89 – 118.

Sarkis, Stephanie. 2017. 11 Warning Signs of Gaslighting. *Psychology Today* https://www.psychologytoday.com/au/blog/here-there-and-everywhere/201701/11-warning-signs-gaslighting. Accessed 14th of January, 2021.

Sawyer, Sarah. 2020. Talk and Thought in *Conceptual Engineering and Conceptual Ethics*, Alexis Burgess, Herman Cappelen, and David Plunkett (eds.), pp. 379 – 395. Oxford: Oxford University Press.

Scharp, Kevin. 2013. *Replacing Truth.* Oxford: Oxford University Press.

Schroeter, Laura. 2012. Bootstrapping Our Way to Samesaying. *Synthese* 189: 177 – 197.

Sherman, Benjamin. 2016. There's No Justice: Why Pursuit of a Virtue is Not the Solution to Epistemic Injustice. *Social Epistemology* 30 (3): 229 – 250.

Simion, Mona and Christopher Kelp. 2019. Conceptual Innovation, Function First. *Nous* DOI: doi: 10.1111/nous.12302

Singer, Peter. 1975. *Animal Liberation.* New York: HarperCollins.

Sloman, Steven, Bradley C. Love, and Woo-Kyoung Ahn. 1998. Feature Centrality and Conceptual Coherence. Cognitive Science 22 (2): 189 – 228.

Smith, E. E., and D. L. Medin. 1981. *Categories and Concepts.* Cambridge, MA: Harvard University Press.

Spear, Andrew D. 2018. Gaslighting, Confabulation, and Epistemic Innocence. *Topoi* 39 (1): 229 – 241.

Spear, Andrew D. Forthcoming. Epistemic Dimensions of Gaslighting: Peer-Disagreement, Self-Trust, and Epistemic Injustice'. *Inquiry* DOI: 10.1080/0020174X.2019.1610051.

Spencer, Quayshawn. 2019. A More Radical Solution to the Race Problem. *Aristotelian Society Supplementary Volume* 93 (1): 25 – 48.

Sperber, D. and D. Wilson. 1998. The Mapping Between the Mental and the Public Lexicon. In P. Carruthers and J. Boucher (eds.), *Language and Thought: Interdisciplinary Themes.* Cambridge: Cambridge University Press.

Stark, Cynthia A. 2019. Gaslighting, Misogyny, and Psychological Oppression. *The Monist* 102: 221 – 235.

Sterken, Rachel. 2020. Linguistic Interventions and Transformative Communicative Disruptions in *Conceptual Ethics and Conceptual Engineering,* Alexis Burgess, Herman Cappelen, and David Plunkett (eds.). Oxford: Oxford University Press.

Solomon, K. O., D.L. Medin, and E. L. Lynch. 1999. Concepts Do More Than Categorize. *Trends in Cognitive Science* 3: 99 – 105.

Stalnaker, Robert. 2014. *Context.* Oxford: Oxford University Press.

Stanley, Jason. 2015. *How Propaganda Works.* Princeton, NJ: Princeton University Press.

Thagard, Paul. 1989. Explanatory Coherence. *Behavioral and Brain Sciences* 12: 447 – 457.

Thomasson, Amie. 2020. A Pragmatic Method for Normative Conceptual Work,' in *Conceptual Engineering and Conceptual Ethics*, Alexis Burgess, Herman Cappelen, and David Plunkett (eds.), pp. 435 – 458. Oxford: Oxford University Press.

Tilly, Charles. 2002. *Stories, Identities, and Political Change.* Landham, MD: Rowman and Littlefield.

Tirrell, Lynne. 2019. Words Matter: Trump Bears A Responsibility for El Paso. *The Guardian*. https://www.theguardian.com/commentisfree/2019/aug/10/trump-el-paso-shooting-speech-words.

Tomasselo, Michael. 2008. *The Origins of Human Communication.* Cambridge, MA: MIT Press.

Tomasselo, Michael. 2014. The Ultra-Social Animal. *European Journal of Social Psychology* 44 (3): 187 – 194.

Tversky, Amos. 1977. Features of Similarity. *Psychological Review* 84: 327 – 352.

Vicente, Agustín and Fernando Martínez Manrique. 2016. The Big Concepts Paper: A Defence of Hybridism. *British Journal of the Philosophy of Science* 67: 59 – 88.

Weber, Max. 1978. *Economy and Society: An Outline of Interpretive Sociology*. Berkeley, CA: University of California Press.

Wieskopf, D. A. 2009. The Plurality of Concepts. *Synthese* 169: 145 – 173.

Williamson, Timothy. 2000. *Knowledge and its Limits*. Oxford: Oxford University Press.

Wittig, Monique. 1993/1981. One is Not Born a Woman. *The Straight Mind.* New York: Basic Books.

Woomer, Lauren. 2019. Agential Insensitivity and Socially Supported Ignorance. *Episteme* 16 (1): 73 – 91.

Yuval-Davis, Nira. 2006. Intersectionality and Feminist Politics. *European Journal of Women's Studies* 13 (3): 193 – 209.

Zack, Naomi. 1993. *Race and Mixed Race.* Philadelphia: Temple University Press.

Zack, Naomi. 2002. *Philosophy of Science and Race.* New York: Routledge.