



The  
University  
Of  
Sheffield.

# Conceptual Responsibility

Trystan S. Goetze

A thesis submitted in partial fulfilment of the requirements for the degree of  
Doctor of Philosophy.

The University of Sheffield  
Faculty of Arts and Humanities  
Department of Philosophy

August 2018



## Abstract

This thesis concerns our moral and epistemic responsibilities regarding our concepts. I argue that certain concepts can be morally, epistemically, or socially problematic. This is particularly concerning with regard to our concepts of social kinds, which may have both descriptive and evaluative aspects. Being ignorant of certain concepts, or possessing mistaken conceptions, can be problematic for similar reasons, and contributes to various forms of epistemic injustice. I defend an expanded view of a type of epistemic injustice known as ‘hermeneutical injustice’, where widespread conceptual ignorance puts members of marginalized groups at risk of their distinctive and important experiences lacking intelligible interpretations. Together, I call the use of problematic concepts or the ignorance of appropriate concepts ‘conceptual incapacities’. I discuss the conditions under which we may be responsible for our conceptual incapacities on several major theories of responsibility, developing my own account of responsibility in the process, according to which we are responsible for something just in case it was caused by one of our reasons-responsive constitutive psychological traits. However, I argue that regardless of whether we are responsible for something, we may still be required to *take* responsibility for it. Whether or not we are responsible for our conceptual incapacities, we are required to reflect critically upon them in a variety of scenarios that throw our use of those concepts into question. I consider the method of conceptual engineering — the philosophical critique and revision of concepts — as one way we might take responsibility for our concepts, or at least, defer that duty to experts. But, this top-down model of conceptual revision is insufficient. Using a pragmatist model of the social epistemology of morality, I argue that conceptual inquiry is a social endeavour in which we are all required to participate, to some degree.

**Word Count** (including notes, excluding front and end matter): 70,746



## Table of Contents

<b>Abstract</b> .....	iii
<b>Acknowledgements</b> .....	vii
<b>Introduction</b> .....	1
<b>1. Wrongful Concepts</b> .....	11
1.1. Concepts.....	14
1.1.1. A Working Account of Concepts.....	14
1.1.2. Three Theories of Concepts .....	16
1.2. Thick Concepts — Ethical and Social .....	23
1.3. Conceptual Wrongs .....	31
1.4. Objections.....	40
1.4.1. Why Concepts?.....	40
1.4.2. Impossible Demands?.....	41
1.4.3. Thought Police? .....	43
1.5. Summary.....	46
<b>2. Conceptual Ignorance and Epistemic Injustice</b> .....	47
2.1. Conceptual Ignorance and Conceptual Non-competence.....	47
2.2. Structures of Conceptual Ignorance: Hermeneutical Injustice .....	51
2.2.1. Collective Resources and Dissenting Communities.....	55
2.2.2. The Harms of Hermeneutical Injustice .....	59
2.2.3. Global and Local Hermeneutical Economies .....	63
2.3. Wilful Conceptual Ignorance and Non-competence .....	73
2.4. Summary.....	79
<b>3. Responsibility for Conceptual Incapacities</b> .....	81
3.1. Three Senses of Responsibility .....	82
3.2. Voluntarism .....	89
3.3. The Deep Self View .....	95
3.4. The Rational Relations View .....	98
3.5. Psychologism .....	104
3.6. A Hybrid View: Reasonable Psychologism.....	114
3.7. Responsibility for Conceptual Incapacities .....	123
3.8. Summary.....	130
<b>4. Conceptual Answerability</b> .....	131
4.1. Taking Responsibility and Being Responsible .....	132
4.2. Bad Luck and Penumbral Agency .....	137
4.3. Moral and Epistemic Answerability.....	148

4.4. Summary .....	157
<b>5. Taking Responsibility for Our Concepts.....</b>	<b>159</b>
5.1. Occasions to Reflect.....	160
5.2. Philosophical Reflection .....	171
5.3. Conceptual Inquiry .....	183
5.4. Summary .....	197
<b>Conclusion.....</b>	<b>199</b>
<b>Bibliography .....</b>	<b>205</b>

## Acknowledgements

I entered this degree without a clear idea of what I wanted to achieve. However, re-reading my initial proposal, there remains a consistent theme: the ethics of the ways we interpret our experiences. I am grateful to the admissions committees at Sheffield for having faith that my confused thoughts at that early stage might end up producing something worthwhile. I hope I have lived up to that expectation.

I want first of all to say thank you to everyone in the philosophy department at Sheffield. I cannot think of a more supportive environment for writing a philosophy thesis. The two people here who have had the greatest impact on my work are, of course, my supervisors, Miranda Fricker and Paul Faulkner. I thank Miranda for her always encouraging remarks that helped me to build up the main ideas of the project. I thank Paul for applying his critical eye to press me to be more clear and precise on a great many points. For their feedback on my work in both written and spoken forms at Sheffield, I also want to specifically thank Ashley Pennington, Charlie Crerar, Simon Barker, Graham Bex-Priestley, Lewis Brooks, Alexandre Duval, Andreas Bunge, James Lewis, Matt Cull, Rosa Vince, Katharine Jenkins, Luca Barlassina, Steve Laurence, Bob Stern, Chris Hookway, Jenny Saul, and Paniel Reyes Cardenas. I am sorry to anyone whom I have omitted. Any remaining errors or foolishness herein is mine alone.

Outside of Sheffield, one of the most useful places I received feedback on my work was at meetings of the Canadian Philosophical Association (2016, University of Calgary; 2017, Ryerson University; 2018, Université du Québec à Montréal). I want to thank Rebecca Mason and David Collins in particular for their commentaries on my work. I am also grateful to Tracy Llanera and Teresa Allen for inviting me to speak on some of the issues discussed in the thesis at the Social Epistemology Working Group at the University of Connecticut. Thanks also to the publisher of *Hypatia* for granting me permission to re-use material I published in their Winter 2018 issue.

In some small way, my thoughts in this thesis have also been shaped by conversations I had much earlier in my graduate studies at York University and the University of Toronto with Lorraine Code, Doug Hutchinson, Jagdish Hattiangadi, Dan MacArthur, Bob Myers, Mike Giudice, Jim Vernon, Kari Delhi, Njoki Wane, Tanya Titchkosky, Lauren Bialystok, and Eric Bredo. It is also unlikely that I would have pursued philosophy as far as I have without the encouragement of Katie Plaisance and Shannon Dea during my undergraduate studies at the University of Waterloo, and that of my philosophy teacher at Kitchener-Waterloo Collegiate and Vocational School, Phil Allt. I thank all of my teachers for making me who I am.

I wish to thank the Philosophy Program at the Graduate Center of the City University of New York for hosting me as a Visiting Scholar in November–December 2017, where I pieced together the final components of the initial draft. I am thankful to Nathan Haydon for facilitating a pair of visits I made to Waterloo in winter 2018. I also thank Megan Kearney for allowing me to make use of a desk in the Comic Book Embassy while I finished the final draft. I am especially grateful to the University of Sheffield for its generous financial support over the first three years of my Ph.D. I am also grateful to Yonatan Shemmer and his family for hosting me during a campus visit while I was deciding where to take the PhD.

While I value all the friendships and connections I have formed over the course of my graduate studies, I am especially thankful to have spent time with Stephen Bolton, Isela González Vázquez, Ashley Pennington, Josh Black, Phill Nash, Philip Rau, Ahmad Fattah, Holly Longair, Charles Côté-Bouchard, and Shadi Afshar. The love and support of my mom, dad, and brothers has also been essential — and I am sorry not to have made more time to come home to visit! Finally, to Kat: for many reasons, I could not have finished this project without you. Thank you, and love always.

*Trystan S. Goetze*

*Sheffield, August 2018*



*[There is] embedded in the very substance of all our thoughts about the world and about ourselves an inalienable and ineradicable framework of conception, which is not of our own making, but given to us ready-made by society — a whole apparatus of concepts and categories, within which and by means of which all our individual thinking, however original and daring, is compelled to move. This common inherited scheme of conception, which is all around us and comes to us as naturally and unobjectionably as our native air, is none the less imposed upon us and limits our intellectual movements in countless ways — all the more surely and irresistibly because, being inherent in the very language we must use to express the simplest meaning, it is adopted and assimilated before we can so much as begin to think for ourselves at all. This mass of collective representation is, of course, constantly undergoing gradual change, largely due to the critical efforts of individual thinkers, who from time to time succeed in introducing profound modifications. It is different for every age in history, for every well-marked group in the intellectual chart of mankind, and even within such groups, in a minor degree, for every nationality.*

—Francis MacDonald Cornford (1912, 44–45).



## Introduction

In 1947, the German philologist Victor Klemperer published a notebook on the language of the Nazi regime. He refers to the characteristic words and phrases often repeated by Nazi mouthpieces as the *Lingua Tertii Imperii* (LTI), a term mocking the Nazis' fondness for acronyms and potent-sounding foreign phrases. The LTI contains many terms and phrases that, Klemperer observes, fell out of use in the post-war years; but there were others that had a curious staying power. Klemperer warns the reader that these sometimes apparently innocent ways of speaking are insidious:

Words can be like tiny doses of arsenic: they are swallowed unnoticed, appear to have no effect, and then after a little time the toxic reaction sets in after all... The task of making people aware of the poisonous nature of the LTI and warning them of its dangers is, I believe, not just schoolmasterish. If a piece of cutlery belonging to orthodox Jews has become ritually unclean they purify it by burying it in the earth. Many words in common usage during the Nazi period should be committed to a mass grave for a very long time, some for ever.  
(Klemperer 2002, 15-16)

The 'poison' that Klemperer refers to is not the words and phrases of the LTI themselves. The vocabulary of the LTI is rather the delivery mechanism for the toxin, whose effects are principally to change the ways people *think*. His book opens with the example of how Nazi language affected the concept of *heroism*:

What a huge number of concepts and feelings it has corrupted and poisoned! At the so-called evening grammar school organized by the Dresden adult education centre, and in the discussions organized by the *Kulturbund* {Cultural Association} and the *Freie deutsche Jugend* {Free German Youth} I have observed again and again how the young people in all innocence, and despite a sincere effort to fill the gaps and eliminate the errors in their neglected education, cling to Nazi thought processes. They don't realize they are doing it; the remnants of linguistic usage from the preceding epoch confuse and seduce them. We spoke about the meaning of culture, or humanitarianism, of democracy and I had the impression that they were beginning to see the light... and then, it was always just round the corner, someone spoke of some heroic behaviour or other, or of some heroic resistance, or simply heroism per se. As soon as this concept was even touched upon, everything became blurred, and we were adrift once again in the fog of Nazism. And it wasn't only the young men who had just returned from the field or from captivity, and felt they were not receiving sufficient attention, let alone acclaim, no, even young women who had

not seen any military service were thoroughly infatuated with the most dubious notion of heroism. The only thing that was beyond dispute, was that it was impossible to have a proper grasp of the true nature of humanitarianism, culture, and democracy if one endorsed this kind of conception, or to be precise misconception, of heroism. (2, emphasis mine)<sup>1</sup>

The Nazi conception of heroism was built around valorising ‘bravery and foolhardy, death-defying behaviour in some military action or other’ (5), and was always tied to vainglorious and ostentatious public displays of accolade. This emphasis on the prideful, strong, aggressive, and fanatically devoted warrior erases those Klemperer views as the real heroes of the Second World War: the people who were murdered in the concentration camps, and the ‘Aryans’ who quietly shielded Jews and others persecuted by the Nazis.<sup>2</sup> When total war made every civilian into a kind of soldier, the image of the Nazi hero infected all levels of society: ‘military heroism is stored in every factory, in every cellar; children, women, and old people die one and the same heroic death in battle... normally only befitting or attainable by young soldiers in the field’ (4–5).

Klemperer notes that although this view of heroism has always existed in connection with sport and military service, it became one of several ideas that the Nazis repeated incessantly not just in their speeches, but also in the newspapers, books, and entertainment media that they controlled. Their overall aim was — by emphasizing the physical over the mental, the state over the individual, and fanatical loyalty to the ruling party over free thinking — to make the country and its people into a functional machine with no trace of humanity left.<sup>3</sup> The Nazis’ control of the language

---

<sup>1</sup> The phrases in {brace brackets} are the translator’s gloss.

<sup>2</sup> This latter group includes Victor Klemperer’s wife, Eva. The son of a rabbi, his marriage to an ‘Aryan’ woman was all that saved Victor from being sent to the concentration camps on multiple occasions.

<sup>3</sup> Cf. Hannah Arendt: ‘Men insofar as they are more than animal reaction and fulfillment of functions are entirely superfluous to totalitarian regimes. Totalitarianism strives not toward despotic rule over men, but toward a system in which men are superfluous. Total power can be achieved and safeguarded only in a world of conditioned reflexes, of marionettes without the slightest trace of spontaneity’ (Arendt 1976, 457).

people used day-to-day facilitated the control of their thoughts and feelings, becoming 'the most powerful Hitlerian propaganda tool':

Nazism permeated the flesh and blood of the people through single words, idioms, and sentence structures which were imposed on them in a million repetitions and taken on board mechanically and unconsciously... But language does not simply write and think for me, it also increasingly dictates my feelings and governs my entire spiritual being the more unquestioningly and unconsciously I abandon myself to it. (Klemperer 2002, 15)

In order to complete the process of de-Nazification, Klemperer writes, 'it isn't only Nazi actions that have to vanish, but also the Nazi cast of mind, the typical Nazi way of thinking and its breeding-ground: the language of Nazism' (2).

\* \* \*

An important element of Klemperer's discussion of heroism concerns the use and abuse of *concepts* in our habits of thinking. Academic philosophy, of course, is no stranger to critical analysis of concepts. Perhaps the most general and enduring slogan of Anglophone professional philosophy in the last century is that 'the task of philosophy is to clarify our concepts' (Dummett 2010, 13). Isaiah Berlin describes the purpose of philosophy as 'to extricate and bring to light the hidden categories and models in terms of which human beings think', that is to say, 'categories, concepts, models, ways of thinking or acting, and particularly the ways they clash with one another' (Berlin 1999, 10, 11). And though they come from a rather different tradition, the definition of philosophy offered by Gilles Deleuze and Félix Guattari is 'the art of forming, inventing, and fabricating concepts' (Deleuze and Guattari 1994, 2). But it is only very recently that philosophical attention has turned to *ethical* questions surrounding our conceptual practices.

The philosophical process of conceptual clarification, revision, and invention is often treated as a rarefied and arcane activity with little practical relevance. But this, as Alasdair MacIntyre remarks, is obviously mistaken:

Philosophy leaves everything as it is — except concepts. And since to possess a concept involves behaving or being able to behave in certain ways in certain circumstances, to alter concepts, whether by modifying existing concepts or by making new concepts available or destroying old ones, is to alter behavior.  
(MacIntyre 1998, 2)

Similarly, Berlin notes that the concepts people use to understand their experience ‘must deeply affect their lives, not least when they are unconscious; much of the misery and frustration of men is due to the mechanical or unconscious, as well as deliberate, application of models where they do not work’ (Berlin 1999, 10). Other moral philosophers have occasionally noted that changing our ethical concepts can itself be ethically significant. Iris Murdoch, for example, writes that thinking about other people using inappropriate concepts can be morally wrong: ‘a smart set of concepts may be a most efficient instrument of corruption’ (Murdoch 1970, 32), and by changing the way we represent others, we come to see them not only more accurately, but more justly. Bernard Williams observes that ‘the modern world is marked by a peculiar level of reflectiveness’ (Williams 1985, 163), which puts a significant range of our ethical knowledge in danger. For the more deeply we reflect upon the ethical concepts we have inherited from our cultural outlook, the greater the risk that we will find these concepts arbitrary or not obviously better for securing a good life than any other way of carving up the social world. But because our ethical concepts are ‘action-guiding’, a change in our conceptual repertoire in this domain will have profound effects on how we live.

What is less often appreciated is the upshot of our conceptual practices in other domains. Drawing on Klemperer and Murdoch, Jason Stanley argues that a significant component of a flawed ideology is ‘the concepts it contains (or fails to contain)’ (Stanley 2016, 200). Propaganda, on Stanley’s account, can activate the problematic concepts contained in a flawed ideology in order to influence behaviour, or it can ensure that the population fails to grasp certain concepts that are needed to criticize a flawed ideology. One of Stanley’s examples is the Chinese government’s directions to

university administrators to ensure that their lecturers do not teach subjects such as freedom of the press, civil rights, and historical mistakes of the Communist Party: ‘This is a clear attempt to ensure that students lack crucial political concepts, precisely the ones possession of which would enable them to critique Chinese government policy. It is an attempt to install a flawed ideology in Chinese students by ensuring that they lack crucial *political concepts*’ (203, emphasis mine).

But there need not be such top-down control for problematic concepts to take hold in society and do material harm. Patricia Hill Collins describes several concepts of *social* kinds that delineate the typical ways in which black women are represented in mainstream U.S. American culture: mummies, quiet and faithful domestic servants; matriarchs, unfeminine and domineering single working mothers; welfare mothers, lazy and poor role models for their children; and jezebels, hypersexual young women who exist to satisfy male desire (Collins 2009, 80–90). She calls these concepts ‘controlling images’ because they constrain the ways black women can live by ‘mak[ing] racism, poverty, and other forms of social injustice appear to be natural, normal, and inevitable parts of everyday life’ (77). Resisting black women’s oppression, Collins argues, requires activists to resist these social concepts by propagating different images of black women’s lives.<sup>4</sup>

Going further, John McDowell claims that all aspects of empirical experience are subject to (potentially radical) change in the light of conceptual revision. In a picture of experience that draws on Kant and Hegel, McDowell argues that all beliefs are ultimately founded in conceptually structured perceptions: ‘We should understand what Kant calls “intuition” — experiential intake — not as a bare getting of an extra-

---

<sup>4</sup> Of course, these images are sometimes exploited by those in power. As Stanley remarks: ‘the efforts of propagandists to tie the language of poverty and aid to the supposed inferiority of American citizens of African descent have made democratic deliberation about how to handle poverty impossible’ (Stanley 2016, 6); ‘the failure of fit between white belief and Black reality appears inconsistent with the possibility of democratic deliberation’ (16).

conceptual Given, but as a kind of occurrence or state that already has conceptual content' (McDowell 1996, 9). This 'passive' deployment of our concepts is what enables experience to stand in relations of justification, and at the same time leaves our experience open to reflective criticism: 'it must be possible to decide whether or not to judge that things are as one's experience represents things to be. How one's experience represents things is not under one's control, but it is up to one whether one accepts the appearance or rejects it' (11). One way our beliefs can go wrong is where our *concepts of all domains* — moral, social, empirical, and so on — are themselves in error. Thus, we all have what McDowell calls a 'standing obligation' to reflect critically on the concepts we use in both our active thinking and our passive perception (12).

That we sometimes succeed in effecting change in our concepts, and that this change has ethical, political, and epistemic import, suggests that our conceptual activities are open to *appraisal*, be it in terms of moral value or some other domain. So too, our active participation in these practices — as philosophers or as engaged members of lay publics — is open to such appraisal. Certainly, the concepts discussed by Klemperer as components of the LTI call out for censure. Moreover, the Nazi politicians and intellectuals who pushed them, perhaps the members of the public who uncritically absorbed and used them, and certainly those who would bring these ways of thinking back today, call out for disapprobation. This suggests that we may be *responsible* for our conceptual practices, that we may be held *accountable* for using certain concepts, and that, when we are held in the grip of some problematic ways of thinking, we are required to *take* responsibility for our concepts. This thesis is, to my knowledge, the first extended philosophical treatment to take these issues of *conceptual responsibility* seriously.



In the first chapter, I examine a number of *ways in which our concepts can go wrong*. I start by outlining a working account of what concepts are — i.e. the constituents of thoughts — which I argue is neutral between three major theories of the nature of concepts. I then discuss a particular range of concepts that have both descriptive and evaluative aspects. These *thick concepts* are at the heart of much of our moral practices and understanding. But concepts of other kinds can also be thick, such as the concepts corresponding to derogatory social categories. I argue that the use of such concepts can be wrongful not only because they issue in harmful action, but also because they may partially constitute vicious character traits, and the attitudes they construct wrongfully dehumanize the people so represented. I also respond to several objections to the moral evaluation of concepts and thoughts.

In the second chapter, I discuss the issue of *conceptual ignorance* and its connection to epistemic injustice. Lacking certain concepts, or possessing mistaken conceptions, can produce similar harmful effects as the use of problematic concepts. Together, I call the use of problematic concepts and conceptual ignorance *conceptual incapacities*. But the right concepts are not always immediately available. One effect of the oppression or marginalization of particular social groups is that the concepts they have developed to understand their distinctive and important social experiences may not receive widespread uptake. This puts them at risk of *hermeneutical injustice*, a form of epistemic injustice where one's social experience lacks intelligibility because of one's aforementioned marginalization. I develop a view of hermeneutical injustice in response to critical discussion in the literature, explicitly making room for social groups to develop their own concepts where prevailing interpretations run against them. I also propose a model of what I call the *hermeneutical economy*: the different conceptual repertoires that are available to different social groups. I then raise the question: could we be culpable for our conceptual ignorance? While some have argued that we can, I show that, on at least one prominent view of responsibility, very few

cases of conceptual ignorance would meet the conditions for responsibility, and so *a fortiori*, that we would hardly ever be culpable for them.<sup>5</sup>

In the third chapter, I turn to the question of whether we are *responsible* for our conceptual incapacities in detail. I first distinguish between three senses of ‘responsibility’ that are relevant here. In the *attributability* sense, the agent is responsible for something where it can be connected to her agency such that she is open to moral appraisal for it. In the *accountability* sense, the agent is responsible for something where she is open to praise, reward, blame, punishment, or other kinds of third-personal response. In the *answerability* sense, the agent is responsible for something where she is required to take responsibility for it, i.e. to apologize for, justify, acknowledge, make amends for, or otherwise respond to it in a significant way. I argue that attributability is the fundamental sense in most cases, and that culpable fault requires the coincidence of the three. I then go over several major theories of responsibility-attribution, remarking on their answers as to whether we are responsible for our conceptual incapacities, but rejecting each in turn for failing to accommodate our intuitions. I then present my own theory, which I call *reasonable psychologism*, and go over several cases showing the conditions for being responsible for our conceptual incapacities.

In the fourth chapter, I argue that even when we are not responsible for our conceptual incapacities, in the sense that they reflect badly upon us, we are still morally and epistemically required to *answer* for them, i.e., to take responsibility for them. I argue that taking responsibility, in the relevant sense, is not a matter of *deciding* to take on certain obligations, but rather of *fulfilling* obligations one has by virtue of some connection to an event. Often, the relevant events are attributable to the agent, but in cases such as bad moral luck, the agent is required to take responsibility for

---

<sup>5</sup> Part of this chapter has been published in *Hypatia* (see Goetze 2018). Permission has been granted by the publisher to reuse that material here, in lightly modified form.

things that she is not responsible for in the attributability sense. What makes us answerable, despite not being responsible in the core sense, for these events, is their connection to morally important aspects of our identity. Drawing an analogy between the moral sense of answerability that I have defended here, and an epistemic sense in which we are required to answer for our beliefs, I argue that with regard to our conceptual incapacities, whether we are culpable for them or not, we are morally and epistemically required to take responsibility for them by engaging in critical reflection.

In the fifth chapter, I outline how, as individuals and as members of communities, we may engage in such *critical reflection*. I start by going over several occasions where we should be spurred to reflect in this way, such as when we experience feelings of alienation from certain concepts, when we encounter people who refuse to use certain concepts, when we look closely at the history of our concepts, and when we experience hermeneutical injustice. Once we set off to critically reflect on our concepts, however, we may find that we lose confidence in them. Sometimes, it is appropriate merely to set those concepts aside, for use only in historical interpretation. But in many other cases, it is important that we find some way to *replace or revise* the concepts under scrutiny. I consider one burgeoning approach in philosophy, known as *conceptual engineering*, where philosophers use specialized methods to clarify, critique, and revise our concepts. But, I argue that this top-down imposition of concepts developed in a theoretical context, and our deference to theorists who have done so, cannot, on its own, be all there is to taking responsibility for our concepts. In addition, we need a bottom-up account of how individuals and communities effect conceptual change. Drawing on a pragmatist account of the social epistemology of morality, I present a model of *conceptual inquiry*, whereby individuals and groups, lay publics and experts, criticise and change their concepts. Taking responsibility for our concepts requires us to participate in these inquiries in good faith.

I conclude the thesis with a summary of the major points, and speculation on the consequences of the views defended herein — in other words, a programme for future work on conceptual responsibility.

## 1. Wrongful Concepts

Much of philosophy is concerned with the critical analysis and improvement of our concepts. Sometimes, this work has purely theoretical goals. For instance, explicating the content of a scientific (e.g. SPECIES) or metaphysical (e.g. TIME) concept could reveal confusion in our thinking that a critically revised concept would rectify, in turn improving our theories.<sup>6</sup> However, sometimes we have practical or political goals in mind when we undertake to analyse or improve a concept — and changing social conditions may shift our concepts in turn. As Sally Haslanger observes, ‘our concepts and our social practices are deeply intertwined. Concepts not only enable us to describe but also help structure social practices, and our evolving practices affect our concepts’ (Haslanger 2012b, 368). For example, a queer analysis and critique of traditional sex and gender concepts (e.g. MALE, FEMALE; MAN, WOMAN) might be aimed at showing their inadequacy for covering all the ways in which people identify themselves and conduct their lives, proposing instead a wider range of sex and gender concepts or the elimination of these categories altogether. This sort of project may be aimed at the social and political goal of achieving rights and recognition for queer folks.

That we can have practical and political goals in mind when scrutinizing our concepts raises the possibility that the concepts we have, and the ways in which we use them, are open to moral appraisal. For instance, it may be *morally wrong* to use certain concepts in a particular way. As Iris Murdoch observes, ‘certain ways of describing people can be corrupting and wrong. A smart set of concepts may be a most efficient instrument of corruption’ (Murdoch 1970, 32). Murdoch describes an example of a mother, M, who finds her daughter-in-law, D, to be distasteful in various ways:

M finds D quite a good-hearted girl, but while not exactly common yet certainly unpolished and lacking in dignity and refinement. D is inclined to be pert and

---

<sup>6</sup> Throughout, I follow the convention of labelling concepts in SMALL CAPS. Where authors quoted use a different convention for labelling concepts (e.g. *italics* or ‘quotes’), I have silently changed their formatting.

familiar, insufficiently ceremonious, brusque, sometimes positively rude, always tiresomely juvenile. M does not like D's accent or the way D dresses. M feels that her son has married beneath him. (16–17)

However, M is a reasonably intelligent and good-willed person, and later subjects her initial impression of D to greater scrutiny. M is self-conscious enough to realize that her upbringing in a conservative and elitist social environment, not to mention her own jealousy at D's relative freedom, may have coloured the way in which she had characterized her daughter-in-law. Upon further reflection, M realizes that D is not distasteful at all, but rather that M's own value system was snobbish and old-fashioned. Murdoch emphasizes that this change in perspective comes about because of a change in the concepts M uses: 'D is discovered to be not *vulgar* but refreshingly *simple*, not *undignified* but *spontaneous*, not *noisy* but *gay*, not tiresomely *juvenile* but delightfully *youthful*, and so on' (17 my emphasis). In coming to characterize D in different terms — that is to say, by using different *concepts* — M corrects a moral error: she is attempting 'not just to see D accurately but to see her justly or lovingly' (23).

Murdoch's example is a relatively tame one. It may well be unfair to D to think of her as, say, 'undignified' rather than 'spontaneous', and this may well sour familial relations between M and her son and in-laws should they ever find out what M really thinks of D. But it is easy to find cases where more is at stake. Consider Frantz Fanon's account of his encounters with racism. He says that before coming to France he thought of himself as just one man among others, but his racialization made the continuation of this self-identification impossible: white folks had categorized him as a 'Negro', which forced him to reframe how he thought of himself in consideration of how white folks thought of him. Rather than simply indicating his ancestry or skin colour, the category 'Negro' has, as Fanon puts it, 'historicity', that is to say, a history of use and associations tied to social practices. Fanon writes that his social identity was in some sense wrested away from him by this categorization:

My body was given back to me, sprawled out, distorted, recolored, clad in mourning on that white winter day. The Negro is an animal, the Negro is bad,

the Negro is mean, the Negro is ugly... I sit down at the fire and I become aware of my uniform. I had not seen it. It is indeed ugly. I stop there, for who can tell me what beauty is? (Fanon 2008, 86)

He describes himself as being ‘imprisoned’ by this racist characterization, because everywhere he goes he finds white people calling him a ‘Negro’ (or worse), making assumptions about his character or intelligence on the basis of his race, even fearing that because of his ancestry he may be dangerous. The case could be made that the white people in Fanon’s story simply have a set of racist beliefs about people of African descent. But the emphasis Fanon lays on how white folks think of Fanon as a ‘Negro’ and not as ‘a man, nothing but a man’ (85) is suggestive. It is possible that their prejudice is operating at a cognitively prior stage — at the level of their *concepts*. And this would suggest not only that the use of certain concepts can be morally wrong, but also that the use of such concepts may be implicated in systemic oppression.

In this chapter, I explain how certain concepts may be morally wrong to use in thought. Of course, the nature of concepts is an unsettled question in the philosophy of mind, language, and the cognitive sciences. Settling the metaphysics of concepts is not my project here, nor do I aim to endorse any particular *theory of concepts*. Instead, in the first section, I briefly develop a working account that is consistent (or, at least, not problematically inconsistent) with three major kinds of theories of concepts. In the second section, I examine a particularly compelling case where the use of certain concepts is morally wrongful: that of so-called ‘thick’ concepts, which have both evaluative and descriptive dimensions. While the paradigm instances of thick concepts are usually drawn from ethics, there are also cases of thick concepts of *social* kinds. The latter are of particular interest, given their connection to discriminatory practices and oppressive social institutions. In the third section, I describe three ways in which using certain concepts can be wrongful or otherwise problematic, drawing on consequentialist, aretaic, deontological, and social-ontological considerations. In the fourth section, I consider several objections to the claim that using certain concepts

could be morally wrongful. The first is the suggestion that we should focus on morally wrongful attitudes, rather than concepts. The second is that the notion that certain concepts may be morally wrong to use may generate impossible demands, or lead to a self-effacing theory. The third is that by extending moral appraisal to the ‘inner world’, I am opening the way to the danger of thought policing.

### 1.1. Concepts

In this section, I outline a working account of concepts that will be in the background for the remainder of the thesis. I outline three major theories of concepts and argue that my working account is consistent, or at least not problematically inconsistent, with each of them.

#### 1.1.1. A Working Account of Concepts

Our concepts construct our propositional attitudes. To possess a concept of X, the subject must be able to entertain and acquire propositional attitudes containing that concept. Minimally, this requires that the subject be able to identify things in the world as X’s or to think about X’s *as such*.<sup>7</sup> In order to possess the concept SQUARE, for example, the subject must be able to identify particular objects as squares and to acquire attitudes such as beliefs that those objects are squares. This involves thinking about those objects *as squares*, and not under some other category, say, shapes.

Usually, however, using a concept is more than just attaching a label to a class of objects. In order to *master* the use of a concept, the subject must have some beliefs

---

<sup>7</sup> Cf. Édouard Machery’s definition of the way ‘concept’ is used in philosophy: ‘Having a concept of x is being able to have propositional attitudes about x as x’ (Machery 2009, 32). Machery argues that philosophers and psychologists are actually interested in different things when they talk about concepts, with philosophers taking an interest in how our thoughts have semantic content, and psychologists interested in how our brains process information (cf. fn. 10, 14 below; and see Machery 2009, 7–51, esp. 34ff.). I am not convinced that the two theoretical projects are as distinct as Machery claims, since the mental processes psychologists discuss when theorizing about concepts typically include the sorts of processes that produce propositional attitudes, such as reasoning and categorizing.



regarding the upshot of identifying objects under it.<sup>8</sup> This notion of concept mastery is vague; just how much the subject must believe before having mastered the concept is unclear. It is also unclear whether certain kinds of beliefs are essential to concept mastery — some plausible instances include beliefs about the typical properties of the concept's objects, predictions of the causal interactions explaining the behaviour of the concept's objects, or knowledge of the inferences that one is entitled to make on the basis of committing to a proposition containing the concept. It may be that how many and which beliefs the subject must have to master a concept varies from case to case. But generally speaking, I take it that the notion of having some significant range of beliefs regarding the upshot of making these distinctions is clear enough. For example, mastery of the concept SQUARE may involve knowing that when one identifies an object as a square, one may conclude that it has four interior right angles. The subject may *possess* a concept before she comes to *master* it: children, for example, learn to identify squares long before they learn that all squares have four interior right angles. The subject might also fail to master a concept by acquiring the wrong beliefs regarding the upshot of using that concept. A child might mistakenly believe that squares are not rectangles, for example, and thus fail to know that she is entitled to conclude of any square that it is a rectangle.

Different concepts may be used to think about the same object. I can frame my thoughts about some particular square in terms of the concept SQUARE or the concept RHOMBUS, for instance, depending on which aspects are most relevant in the present context. Sometimes, however, it seems that we have different *variations on the same concept* that attach to the same object: the referent remains fixed, but the beliefs that constitute mastery differ. Following Hilary Putnam, I will mark these variations by

---

<sup>8</sup> I say 'beliefs' and not 'knowledge' here because, as we shall see, mastering some concepts may require believing some falsehoods.

distinguishing between concepts and *conceptions*. Putnam illustrates using differences in the use of the concept of temperature over the history of science:

When we translate a word as, say, *temperature* we equate the reference and, to the extent that we stick to our translation, the sense of the translated expression with that of our own term 'temperature', at least as we use it in that context... In this sense we equate the 'concept' in question with our own 'concept' of temperature. But so doing is compatible with the fact that the seventeenth-century scientists, or whoever, may have had a different *conception* of temperature, that is a different set of beliefs about it and its nature than we do, different 'images of knowledge', and different ultimate beliefs about many other matters as well. (Putnam 1981, 117)

When it comes to squares, perhaps there is, on the one hand, a common-sense conception of squares, mastery of which includes the belief that squares are shapes with four sides of equal length, and, on the other hand, a mathematical conception of squares, mastery of which includes the belief that squares are shapes with four sides of equal length *and* four interior right angles. The former suffices for most everyday geometry, but the latter is more precise (and thus, more suitable for doing mathematics) in that it allows for a distinction between, for example, squares and rhombuses, where both are equilateral but only the former are equiangular.

### 1.1.2. Three Theories of Concepts

It is important before going on that we check to see that my working account of concepts, concept mastery, and conceptions is consistent with the major theories of concepts. For without doing so, my project runs the risk of floating in a philosophical aether, without sufficient grounding to have much practical upshot. The three theories of concepts that I will examine are: (i) mental representations, (ii) cognitive abilities, and (iii) Fregean senses.<sup>9</sup>

On the first kind of view, concepts are *mental representations*. These views hold that the mind is a symbolic system of psychological entities, such as beliefs, desires,

---

<sup>9</sup> In this subsection I have benefitted greatly from Stephen Laurence and Eric Margolis's helpful overviews of the various theories of concepts; see Margolis and Laurence (2014, 2007), Laurence and Margolis (1999). I thank Steve Laurence for additional discussion.

and other propositional attitudes. These psychological entities are in turn composed of more fundamental entities. The basic unit of mental representations is *concepts*. The concept SQUARE, on this kind of view, would be a psychological entity with some kind of content that allows it to represent squares and to figure in propositional attitudes about squares. There is considerable debate, amongst adherents of mental representations views, over what conceptual content consists in. One view holds that concepts are *atomic*, that is to say, they have no internal structure and their semantic content is determined not by their relations to other concepts or other representations but to the world (e.g. Fodor 1998). Among those who hold that concepts get their content from some internal structure, various possibilities have been advanced. Concepts might have a structure analogous to *definitions*, that is to say, containing further representations that provide the necessary and sufficient conditions for applying the concept (this “classical” theory has a long history in Western philosophy, with roots in, e.g. Plato 2006; Locke 1975; Hume 1978; more recent defenders of this sort of view include Jackendoff 1983; Pinker 1989). Concepts might be structured as *prototypes*, that is to say, statistical information about the likely properties of the objects they pick out (e.g. Rosch and Mervis 1975). Concepts might be structured as *theories*, that is to say, information that provides an explanation of the phenomena they pick out (e.g. Carey 1985). Concepts might be structured as *exemplars*, that is to say, information about the properties of one or more paradigmatic members of the category picked out (e.g. E. E. Smith and Medin 1981). Pluralist views, combining aspects of some or all of these forms of mental representations in their accounts of concepts, have also been advanced (Laurence and Margolis 1999; Weiskopf 2009; Machery 2009).<sup>10</sup>

---

<sup>10</sup> Strictly speaking, Machery (2009) argues that psychologists should eliminate their use of the term ‘concept’ and refer only to the more precise entities of prototypes, exemplars, and theories, and that philosophers are interested in a rather different sort of project.

On the second kind of view, concepts are identified with a kind of *cognitive ability*. This account of concepts stems from scepticism, ultimately due to a reading of Wittgenstein, about mental representations. Following this line of thought, Michael Dummett argues that mental representations will not provide a satisfactory account of how concepts have semantic content, because the content of those mental representations will also stand in need of explanation.<sup>11</sup> Instead, Dummett says we should account for concepts in terms of a subject's ability to competently use their language to correctly identify objects in the world:

What is it to grasp the concept SQUARE, say? At the very least, it is to be able to discriminate between things that are square and those that are not. Such an ability can be ascribed only to one who will, on occasion, treat square things differently from things that are not square; one way, among many other possible ways, of doing this is to apply the word 'square' to square things and not to others. And it can only be by reference to some such use of the word 'square', or at least of some knowledge *about* the word 'square' which would warrant such a use of it, that we can explain what it is to associate the concept SQUARE with that word. An ability to use the word in such a way, or a suitable piece of knowledge about the word, would, by itself, *suffice* as a manifestation of a grasp of the concept. (Dummett 1996, 98)<sup>12</sup>

On this account, possessing the concept of X is having the ability to correctly use a word referring to X in one's language — when one sees an X, one would understand that it is correct to say 'There is an X', for example. But competence with the use of a word also involves understanding the consequences of making such identifications. As Dummett remarks elsewhere, 'Learning to use a statement of a given form involves, then, learning two things: the conditions under which one is justified in making the statement; and what constitutes acceptance of [that statement], i.e. the consequences of accepting it' (Dummett 1993, 453). These consequences go well beyond just the correct use of a word; among other things, they include the sorts of conclusions we may draw from the fact that the word applies. Robert Brandom has taken such

---

<sup>11</sup> See Laurence and Margolis (1997) for a reply to this kind of argument.

<sup>12</sup> Cf. Kenny (2010, 105–6): 'A sufficient, but not a necessary condition for a person to possess the concept of F is that she shall have mastered the use of a word for "F" in some language'.

inferential consequences of using certain words as the basis for his ‘pragmatics-first’ account of meaning. On Brandom’s view, grasping a concept ‘is mastering its *inferential* use: knowing (in the practical sense of being able to distinguish, a kind of knowing *how*) what else one would be committing oneself to by applying the concept, what would entitle one to do so, and what would preclude such entitlement’ (Brandom 2009, 11).

On the third kind of view, concepts are identified with *abstract entities*. The most significant exponent of this view is Christopher Peacocke, who argues that mental representations are too fine-grained and idiosyncratic to give us an account of the semantic role of concepts. On his view, concepts are Fregean senses, individuated on the following basis:

Concepts C and D are distinct if and only if there are two complete propositional contents that differ at most in that one contains C substituted in one or more places for D, and one of which is potentially informative while the other is not. (Peacocke 1992, 2)

That is to say, on Peacocke’s view C and D are distinct if the proposition ‘C is a D’ is informative rather than trivial. For example, the proposition ‘Margaret Atwood is Margaret Atwood’ is trivial; it tells us nothing we did not already know. The proposition ‘Margaret Atwood is Bart Gerrard’ is informative; it tells us that the Canadian author and an obscure political cartoonist from a radical magazine in the 1970s are in fact the same person.<sup>13</sup> On Peacocke’s account, this difference is due to each concept being identified with the semantic content of a term that presents its referent in a particular way. Possessing a concept of X involves being able to correctly identify X’s in the world and, sometimes, being disposed to find certain inferences compelling. Exactly how this shakes out depends upon our analysis of the possession conditions for the concept of X, which on Peacocke’s view requires very complex and precise work to identify the relevant perceptual stimuli and facts about the world

---

<sup>13</sup> See Gzowski (1978).

indicating that the concept applies, as well as the various inferential and other consequences the thinker must recognize.

How does my working account of concepts stack up against these three kinds of theories? Recall that my account has three parts: (i) I specified that concepts construct attitudes and represent their objects in a particular way, (ii) I added that mastering a concept requires having a range of beliefs about the objects of that concept, and (iii) I noted that different collections of beliefs about the objects of some concept may amount to differing conceptions. Let us take these up in turn. First, all three views hold that concepts construct propositional attitudes, though they differ on the specific ontology. This role is obvious on the Fregean senses view, where concepts are the abstract entities that are combined to produce the propositions to which the thinker may acquire attitudes. Mental representations views do not necessarily identify concepts with the constituents of propositional attitudes, but they are all in agreement that the mind at least employs concepts in processes that construct attitudes, such as deducing and categorizing.<sup>14</sup> On the cognitive abilities view, theorizing about psychological entities is side-stepped in favour of concentrating on semantics and pragmatics; it is less important to say what a concept *is* than it is to specify, in terms of language use, what someone with that concept can *do*. But again, on this view concepts are defined (partly) in terms of one's ability to acquire propositional attitudes about the objects of those concepts, as indicated by one's understanding of the conditions for using the corresponding words in an assertion or inference.

Second, each of the three kinds of theories provides a way to capture of the idea of concept mastery, in each case filling in some of the sketchier details regarding which

---

<sup>14</sup> Cf. Machery's general definition of 'concept' as it is used in psychology: A concept of *x* is a body of knowledge [i.e. a prototype, exemplar, or theory] 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 when these processes result in judgements about *x*' (Machery 2009, 12).

beliefs are required for mastery. On non-atomic mental representations views, mastering a concept means having stored sufficient information — in the form of definitions, prototypes, exemplars, or theories — in one’s long-term memory about the objects of that concept. These kinds of information are not beliefs *per se*, but they serve the appropriate function in filling out the details of the notion of concept mastery. On the atomic mental representations view, the kinds of information that the thinker must have in order to master a concept could not be semantic components of the concept itself, though they could be some range of associated beliefs or non-semantic components of the concept, such as the procedure by which the mind makes rapid identifications of objects that fall under that concept’s application. On the cognitive abilities view, to master a concept one must know not just the application conditions for a corresponding word, but also the implications of using it in terms of ‘the inferential powers of the statement and anything that counts as acting on the truth of that statement’ (Dummett 1993, 453). On the Fregean senses view, the nature of a concept just is specified by the knowledge and dispositions required to have mastered the concept; merely being able to use the concept in thoughts, without having mastered it, fulfils the much weaker requirements for attributing attitudes using the concept to a thinker (Peacocke 1992, 29–30).

Third, where we can make sense of the idea of concept mastery, we can make sense of the idea of differing conceptions.<sup>15</sup> Both of these ideas concern the knowledge, beliefs, or information that one must have in addition to knowing how to apply the concept to its objects, that is to say, knowing what the concept refers to. On the atomic

---

<sup>15</sup> In some cases differing ‘conceptions’ will turn out to be different concepts, so speaking of differing ‘conceptions’ will be just a convenient turn of phrase emphasizing that the concepts in question are closely related. Some have claimed that where a dispute is over which *concept* we should use, rather than over which beliefs we should have, one interlocutor has just changed the subject. But I think that in the cases I am interested in, this worry does not obviously apply. See Prinzing (2017) for discussion of this problem.

representations view, different conceptions of X will just be different sets of beliefs about X's that employ the concept of X, or will have differences in their non-semantic content specifying slightly different application conditions. On compositional representations views, different conceptions of X will be concepts of X that vary to some degree in the relevant information about what X's are like. On the cognitive abilities view, different conceptions of X will be differences in belief regarding the consequences of asserting that such-and-such is an X. On the Fregean senses view, we may be forced to say that different conceptions are just different concepts with the same referent, since on this account concepts are individuated in part by specifying the beliefs that a thinker must have in order to possess a concept. But where differences in these beliefs do not quite point to a difference in the sense of the term, we may be able to specify a range of beliefs about the object of the concept in addition to those required for concept possession; different thinkers may have different sets of beliefs, and so different thinkers may have differing conceptions.

So, my working account of concepts is not inconsistent with the three major clusters of theories of concepts that I have summarized. Before moving on to discuss the ways in which it can be wrongful to use a concept, one further feature is worth emphasizing. Since, as we saw, one must have the concept of X to think about X's as X's, the concepts we have constrain the thoughts we can think. And, Alexis Burgess and David Plunkett write,

[This] point isn't merely doxastic. 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 constrains what we can hope to accomplish in the world. Representation enables action, from the most sophisticated scientific research, to the most mundane household task. (Burgess and Plunkett 2013a, 1096–97)

I think it is too strong to say that our conceptual repertoire *determines* the attitudes we can acquire — this would suggest that we are simply stuck with whatever attitudes our concepts allow us to acquire, possibly implying that conceptual change is impossible,



which Burgess and Plunkett would deny. The idea is rather that the ways in which we can think about objects, people, institutions, theoretical entities, and ourselves — the concepts and conceptions that we possess and have mastered — drives our thinking, and thus our actions, down particular paths. And these paths are not always satisfactory: as I argue in the rest of this chapter and the remainder of the thesis, the thoughts and actions that our concepts push us towards may be morally wrong and epistemically problematic.

### 1.2. Thick Concepts — Ethical and Social

To discover the ways in which our concepts may be morally wrong, it will be instructive to direct our attention at a range of concepts that are already involved in our moral practices.<sup>16</sup> Recall the concepts Murdoch refers to in her example of the mother and her daughter-in-law. Murdoch describes the morally significant shift in M's thinking as a change in 'normative-descriptive words' that are embedded in a historical and cultural context (Murdoch 1970, 31–32). As we would now put it, following Bernard Williams's development of this idea, the relevant concepts are 'thick'. That is to say, rather than being 'thinly' evaluative (as the concepts GOOD and BAD, RIGHT and WRONG are supposed to be) or thinly descriptive (as we take concepts such as RED, SQUARE, or CHAIR to be), thick concepts have both descriptive and evaluative aspects. As Bernard Williams puts it, such concepts are action-guiding: 'If a concept of this kind applies, this often provides someone with a reason for action'; but, 'at the same time, their application is guided by the world' (Williams 1985, 140–41).<sup>17</sup> Mastering a thick ethical concept requires learning which situations, objects, behaviours, and so forth call for its application; but at the same time, to apply such a concept is to make a moral

---

<sup>16</sup> In the following I have benefitted from overviews of this topic by Simon Kirchin (2013), Debbie Roberts (2013), and Pekka Väyrynen (2017).

<sup>17</sup> Williams credits the main insights of his account of thick ethical concepts to remarks made by Murdoch and Philippa Foot in a seminar in the 1950s (Williams 1985, 218 n. 7).

evaluation of those situations, objects, character traits, and so forth, and such evaluations are relevant to how we act. For example, suppose I call the Prime Minister a cruel lying coward. The application of those thick concepts (CRUELTY, LIAR, COWARD) provides (*pro tanto*) reasons to take morally significant actions — e.g. protesting his policies and refusing to vote for him or his party’s candidates — and, if he hears that I think of him as such, this gives him a (*pro tanto*) reason to defend himself or to change his behaviour. At the same time, my application of those thick concepts to the Prime Minister relies on facts about his actions and the consequences of his policies.

The internal structure of thick ethical concepts is controversial.<sup>18</sup> Williams rejects what has come to be called the ‘separabilist’ view, according to which the normative and descriptive parts of the concept can be discretely analysed. For the concept to apply, both parts must be realized. For example, when I call the Prime Minister a ‘coward’, COWARD would be analysed as: *c*, and (ethically) bad for being *c*, where *c* is some non-evaluative description of the sort of person we call a coward — say, *X* is *c* when *X* is disposed to run from dangerous situations and backs down from defending her convictions. Williams argues that this kind of account errs in making the world-guidedness of thick ethical concepts *exclusively* descriptive: ‘It follows that, for any concept of this sort, you could produce another that picked out just the same features of the world but worked simply as a descriptive concept, lacking any prescriptive or evaluative force’ (Williams 1985, 141). As a result, the evaluative component thus seems to be an arbitrary rider; there is no reason why the phenomenon in question should have any normative force, or why that force should push us as agents in any particular direction. That I think the Prime Minister is *c* does

---

<sup>18</sup> In addition to the separabilism/inseparabilism debate that I discuss here, Väyrynen (2009) notes that it is not entirely clear how best to theorize the location of the evaluative content associated with thick concepts, an issue that is made particularly salient in the case of refusals to use thick concepts that one objects to on the grounds of those evaluations. I think the broadly Williams-inspired account that I exploit here is a form of what he calls ‘contextualism’, but I will not take up a thorough defence here.

not tell us why that claim makes any ethical demands or evaluations, much less those demands and evaluations which are associated with my thinking that the Prime Minister is a *coward*. I could just as easily describe the Prime Minister as ‘*c*, and neutral (or good!) for being *c*’ — ‘he may run from dangerous situations and back down from defending his convictions, but there is nothing ethically bad about that’. There is nothing about *c* by itself to suggest an ethically bad evaluation is warranted. Moreover, it is unclear why those qualities lumped together under *c* should cluster together in the first place: cowardice and other moral categories seem unlikely to be natural kinds. The separabilist account makes thick ethical concepts seem to be no more than arbitrary preferences for otherwise neutral states of affairs, rather than substantive components of ethical thought.

What makes the descriptive and evaluative aspects of thick ethical concepts analytically inseparable and supplies the substance of their action-guiding aspect, is their embeddedness in a cultural background. That is to say, their ‘world-guided’ aspect is attached not just to objective physical observations, but also to facts about the *social* world. As Williams remarks,

How we ‘go on’ from one application of a concept to another is a function of the kind of interest that the concept represents, and we should not assume that we could see how people ‘go on’ if we did not share the evaluative perspective in which this kind of concept has its point. An insightful observer can indeed come to understand and anticipate the use of the concept without actually sharing the values of the people who use it... But in imaginatively anticipating the use of the concept, the observer also has to grasp imaginatively its evaluative point. He cannot stand quite outside the evaluative interests of the community he is observing, and pick up the concept simply as a device for dividing up in a rather strange way certain neutral features of the world. (Williams 1985, 141–42)

Acquiring and mastering thick ethical concepts happens against the background of practices that have moral significance for the members of a given community.<sup>19</sup> One

---

<sup>19</sup> It is worth noting that Williams makes a distinction between the moral and the ethical. The former refers to a particular normative system in Western society that is marked by notions such as moral obligation and a propensity to eliminate the interference of luck in moral evaluations; the latter concerns a broader range of

has mastered a thick ethical concept only if one has learned to participate in those practices.<sup>20</sup> It is not enough to be able to anticipate the application of a particular word or concept when the relevant descriptions apply; mastering a thick ethical concept requires the agent to grasp the significance of the ethical evaluations that such concepts make and how these are connected to a broader range of practices. For example, if an anthropologist wants to understand the concept COWARD, as it is used in Western culture, she will have to learn not just to anticipate the application of this concept when certain objective conditions obtain — when a person runs away from danger or backs down from defending his convictions — and the accompanying negative evaluation of the person to whom it applies. The anthropologist will also have to appreciate why these activities are given the significance they have in that culture. We value people who stand up to danger and who stand firm in defending their convictions because, for instance, these dispositions are necessary to protect other valuable things — goods, people, places, beliefs — from threat. Since fear can lead us to run away from rather than face danger, or to back down from defending our convictions, the fearful person fails to safeguard these valuable things and warrants our disapproval. The anthropologist must understand this background in order to grasp the concept of a coward.

Against Williams's inseparabilist view, Simon Blackburn counters that disentangling the evaluative from the descriptive is what enables critique of thick ethical concepts. The thought is that criticizing a thick ethical concept is a two-step process. First, we separate out the descriptive and the evaluative. Second, now that the two components have been specified, we ask whether it is good or right to put the two

---

considerations regarding how we should live (Williams 1985). I do not follow this distinction, so I use 'moral' and 'ethical' more or less interchangeably throughout.

<sup>20</sup> Similar remarks about the cultural background of ethical concepts are made by Foot (1954, 2001), Anscombe (1958b).

together. His example is the concept CUTE, as applied to adult women. Of the people who are in the habit of calling women 'cute', Blackburn says:

We may want to say that there is something wrong with them, along the lines of this: they admire and respond excitedly (or perhaps enviously, if they are women) to the non-threatening, infantile, subservient self-presentations that some women consciously or unconsciously adopt. Theirs is a group amongst whom women are successful by presenting themselves as there to be patronized, like pets or babies (which themselves are frequent terms of endearment). And that, we say, is bad. (Blackburn 2013, 123)

Blackburn thinks his critique of CUTE only works because he has disentangled the descriptive parts — non-threatening infantile subservience — from the concept-users' positive evaluation of this behaviour. He claims that his critique consists in saying that it is a *bad* thing for them to see *that* behaviour as *good*. Without disentangling the descriptive and evaluative in CUTE, he claims, there is nowhere for a critique to get purchase; the concept-users can retreat to the idea that they are simply carving up the social world in a different way, which makes different practical demands of them, and that is all there is to say. Making the opposite evaluation — that being cute and valuing cuteness are bad — would merely change the subject: since it is a good thing for a woman to be cute, they might say, any claim to the contrary must be talking about something other than *cuteness*. In order to judge how we should evaluate some range of behaviour (or whatever), Blackburn concludes, we need to disentangle it from any pre-existing evaluations.

Sophie-Grace Chappell replies that, read at face value, the sort of conceptual critique Blackburn's separabilism provides is disappointing: "Describing grown women as cute is bad": is that really *all* we can say about it, on the evaluative side? Simply to say, "using CUTE is bad"... is not much of an argument; it is hard to see why this sort of remark should be expected to impress the cultists of cute' (Chappell 2013, 194). Indeed, when these 'cultists' disentangle CUTE themselves, they will not see a reason to find fault with their positive evaluations of that sort of behaviour: they will arrive at the same description and evaluation, and *approve* of the connection. Our

critique must have more substance than this to be at all persuasive. Indeed, Blackburn himself pitches his critique of CUTE in more substantive terms than flipping from evaluating the relevant behaviour as good to evaluating it as bad. To wit, he makes use of *other* thick concepts to make his point, calling the relevant behaviour ‘non-threatening’, ‘infantile’, ‘subservient’, and ‘there to be patronized’, and making comparisons to the behaviour of babies and pets. As Chappell observes, ‘it is these admittedly thick concepts, and not the allegedly thin concept BAD, that give Blackburn’s argument against CUTE its bite’ (196).<sup>21</sup>

What Blackburn misses is that there *is* plenty more to say to a concept-user’s rebuttal that her culture just carves up the world differently. The critic may, for instance, point out that the concept and its associated practices are inconsistent with other goods. The concept CUTE, if it is as Blackburn describes, could be at odds with a social-political commitment to women’s equality. The idea that timidity, subservience, infantilism, and so on are good qualities for women to have is at cross-purposes with the idea that women may be just as strong, commanding, and mature as men. If the critic shows that valuing both women’s cuteness and women’s equality is inconsistent, then one range of concepts and practices has to go. Critical reflection can remove concepts from their action-guiding use by undermining the cultural background against which continuing to make those distinctions makes sense. We may retain these concepts in a truncated form for interpreting the ways of thinking and living that our ancestors engaged in, but it becomes difficult (and eventually impossible) to return to these ways of thinking and living as the rest of our practices move on. As Williams remarks,

reflection might destroy knowledge, because thick ethical concepts that were used in a less reflective state might be driven from use by reflection... these people once had beliefs of a certain kind, which were in many cases pieces of

---

<sup>21</sup> Chappell (2013) goes on to argue that *all* evaluative concepts, including the allegedly thin concepts GOOD, BAD, RIGHT, and WRONG, are thick concepts. While I am sympathetic to this conclusion, nothing I go on to say depends on this rather strong claim.

knowledge; but now, because after reflection they can no longer use concepts essential to those beliefs, they can no longer form beliefs of that kind'. (Williams 1985, 167)

When we lose thick ethical concepts, we lose with them a range of practices and beliefs that construct our shared social world. Rebuilding conviction in another way of doing things is no easy matter. (I return to the issues of critical reflection and returning to conviction in our concepts in Chapter 5.)

While Williams introduces thick concepts to cover ethical phenomena, Murdoch's understanding of them is much broader. She argues that learning ethical concepts takes place via a similar process to learning various *aesthetic* concepts that are important for artistic appreciation (Murdoch 1970, 1959).<sup>22</sup> This points to a broader range of thick concepts, including, as Simon Kirchin summarizes, 'aesthetic concepts (such as GLAMOROUS, GROTESQUE, SUBLIME), epistemic concepts (such as OBSERVANT, RELIABLE, DULL-WITTED), and many other concepts that might be used to pick out characteristics of people and other things (such as CORNY, FOLKSY, TERRIBLE, TRIUMPHANT, JEJUNE, FASCINATING, IDIOTIC, MESMERIZING, DISAPPOINTING)' (Kirchin 2013, 2). And, recalling the example of Fanon's experience with French racism, we may find that various concepts of social groups and social experiences are not just descriptive, but also carry an evaluative aspect.

Christopher Hom's analysis of pejorative terms suggests precisely this possibility. According to his analysis of slurs, which he calls 'combinatorial externalism' (CE), slurs are thick in a way analogous to the thickness of ethical concepts:

CE is a theory of 'thick' terms, combining both normative and descriptive components into their analyses. For any particular slur, CE holds that its semantic value is a complex, normative property of the form: *ought to be subject to such-and-such discriminatory practices for having such-and-such stereotypical properties all because of belonging to such-and-such group.* (Hom 2012, 394)<sup>23</sup>

---

<sup>22</sup> At times, Murdoch (1970) suggests that the aesthetic and the moral are both dimensions of the same thing — namely, appreciation of the Good.

<sup>23</sup> See also Hom (2008, 431). Hom generalizes CE to apply to all pejorative terms.

Mastering the concept that corresponds to a slur involves knowing how to identify members in the derogated social group, as well as beliefs about the stereotypical properties of members of that group (which particular targets may or may not have). These stereotypical properties will, like the descriptions enabled by thick ethical concepts, already carry an evaluative aspect that, again like thick ethical concepts, prescribes a practical response — to wit, discrimination. Like thick ethical concepts, Hom's account of slurs embeds their meaning in social practices and cultures. Racist slurs, for example, are a component of racist ideologies or racist institutions, and invoke a long history of discriminatory practices. Because of the normative property these concepts point to and their connection to a long history of discrimination on that basis, their corresponding terms have pejorative force that sometimes leaks out and causes offence, even when the speaker is merely mentioning the term rather than using it to derogate.<sup>24</sup>

Recalling Fanon's reflections, the concept NEGRO appears to be a thick social concept in this sense. The French people who have mastered this concept know that it applies to a group of people with certain features or genealogy (namely, dark skin and African ancestry), believe a multitude of stereotypical things about such people (e.g., that they are prone to violence or lacking in intelligence), and judge that for allegedly having such stereotypical properties and belonging to that group, such people should be subject to discriminatory practices. As Fanon remarks, the term has a long history of colonialist myths about and violence towards African people behind it such that even people who attempt to distance themselves from the derogatory use of the term —

“But of course, come in, sir, there is no color prejudice among us. . . . Quite, the Negro is

---

<sup>24</sup> Rae Langton has called this phenomenon ‘leaky quotes’, referring to the practice of using quotation marks in an attempt to distance oneself from the effects of offensive speech: ‘Philosophers expect quotation marks to insulate their contents, like bubble wrap, so that what was harmfully used can be harmlessly mentioned. But there can be leaks, as when the injurious illocutionary potential of a slur seeps through into reported speech’ (Langton 2018, 29).



a man like ourselves” (Fanon 2008, 85) — reinforce rather than undermine the assumptions and derogations that come along with mastery of the concept — ‘It is not because he is black that he is less intelligent than we are. . . . I had a Senegalese buddy in the army who was really clever... Mama, a Negro! . . . Hell, he’s getting mad. . . . Take no notice, sir, he does not know that you are as civilized as we’ (85).<sup>25</sup>

The case of pejoratives shows that thick concepts, particularly in the ethical and social domains, thus have the potential to be morally problematic. Because they are embedded in our cultural practices and make action-guiding evaluations, where those practices are harmful or those evaluations are wrong, the use of these concepts can be morally wrongful. Because of their descriptive aspect, mastery of these concepts requires the thinker to have various beliefs about the objects of these concepts, which may be false or otherwise misleading. In the next section, I discuss these conceptual problems.

### 1.3. Conceptual Wrongs

One obvious way for a concept to be morally wrong to use is that it may produce morally bad actions. Thick concepts provide us with reasons to act. Paradigmatically, the actions in question are morally good, because these concepts are embedded in practices that are considered good or valuable in one’s culture. But the inverse is also possible: a mistaken conception or a concept embedded in a morally problematic practice could provide the agent with putative reasons to take morally bad actions. Consider the concept of chastity. As a thick ethical concept, CHASTITY makes more than a neutral description of premarital celibacy: the notion is embedded in a series of cultural practices that moralise human sexuality, valorise celibacy (especially that of young women), and reinforce traditional gender roles. Its application provides the agent with putative reasons to take a variety of morally problematic actions, such as

---

<sup>25</sup> The four-dot ellipses are in the original.

lying about or refusing even to discuss sexual intercourse with adolescents in the hopes that they will abstain until after marriage. The harms of such deception and reticence are well-documented — and there is good evidence that these methods are not even successful in preventing teenagers from engaging in sexual activity and fail to protect them from sexually transmitted diseases or unwanted pregnancy.<sup>26</sup> The actions to which users of the concept CHASTITY and related notions are guided are demonstrably morally wrong.

The use of certain thick ethical concepts may also indicate a morally vicious disposition. On John McDowell's influential account, virtues consist in a kind of knowledge about which actions are morally required in various situations: 'that the situation requires a certain sort of behaviour is... [the virtuous person's] reason for behaving in that way, on each of the relevant occasions' (McDowell 1998b, 51). Elsewhere, McDowell explains that 'We are alerted to these demands by acquiring appropriate conceptual capacities' (McDowell 1996, 82).<sup>27</sup> In mastering the relevant thick ethical concepts and conceptions, one acquires ethical knowledge regarding how one is morally required to act in response to various practical situations.<sup>28</sup>

There are various ways to fail to be virtuous on this account. The akratic or enkratic, for example, possess but have not quite mastered the relevant concepts, so

---

<sup>26</sup> The authors of one review of abstinence-only education in the United States conclude thus: 'Although health care is founded on ethical notions of informed consent and free choice, federal abstinence-only programs are inherently coercive, withholding information needed to make informed choices and promoting questionable and inaccurate opinions. Federal funding language promotes a specific moral viewpoint, not a public health approach. Abstinence-only programs are inconsistent with commonly accepted notions of human rights.' (Santelli et al. 2006, 79).

<sup>27</sup> McDowell (1996) uses this Aristotelian account of moral perception (or practical wisdom) as a model for the faculty of understanding in general, but in the passage quoted he is specifically referring to ethical demands regarding what to do, rather than epistemic demands regarding what to believe.

<sup>28</sup> For McDowell, combining a suggestion by Aristotle (1925, bk. 1.3, 5.10) and Wittgenstein's worries about concepts that appear to specify rules for behaviour (Wittgenstein 1978, sec. 185), this knowledge cannot be systematically codified, but emerges in practice (McDowell 1998b, 57 ff.).

they find themselves weighing various considerations to determine what they most have reason to do. By contrast, on McDowell's view, when the virtuous person perceives what the situation requires, 'this reason is apprehended, not as outweighing or overriding any reasons for acting in other ways... but as silencing them' (McDowell 1998b, 56).<sup>29</sup> What is distinctive about the *vicious* person, on the Aristotelian view we are considering, is that she is morally misguided to the point that she thinks of herself as being in the right. As Julia Annas writes, 'Aristotle's bad man is like his good man in so far as both display unity of thought and feeling when they act... Aristotle's bad man is someone who has come to have systematically perverted ends, who believes in what he is doing' (Annas 1993, 554). Unlike the akratic or enkratic, whose moral psychologies are pulled in different directions by competing ethical concepts, the vicious person's moral psychology is unified in its direction towards morally bad actions.<sup>30</sup> Since the unified moral psychology of the virtuous person, on this view, is a matter of mastering the right ethical concepts and conceptions, we could explain the vicious person's similarly unified but inverted moral psychology in terms of her having mastered either certain problematic ethical concepts or conceptions, or her being ignorant of the right ethical concepts or conceptions.<sup>31</sup>

---

<sup>29</sup> For criticism of this picture of the virtuous person's moral psychology, see Blackburn (1998). For a defence of McDowell's view, albeit one taking issue with his reading of Aristotle, see Seidman (2005). Nothing in the following turns on whether McDowell is right about this aspect of virtue.

<sup>30</sup> But see Müller (2015) for discussion of an alternative interpretation of Aristotle, according to which the vicious person has a thoroughly unprincipled, disorganized, and capricious moral psychology.

<sup>31</sup> McDowell does not refer to the relevant concepts as 'thick', though it is clear that the concepts he has in mind have both descriptive and evaluative aspects. (Indeed, Williams cites McDowell as another exponent of the view of thick ethical concepts that he develops (Williams 1985, 217–18 n. 7).) Given that McDowell generalizes the Aristotelian view of practical wisdom to cover the use of concepts in empirical experience (cf. fn. 27, above), it is possible that he implicitly accepts a view even stronger than Chappell's (cf. fn. 21, above): to wit, not only are there no thin ethical concepts, there are no thin *descriptive* concepts, either. Again, nothing I say depends on such an account.

For example, on this Aristotelian view, someone who has the vice of cowardice will not merely be disposed to run from fearful situations that he should face — that disposition on its own would be compatible with the fragmented moral psychology of the akratic agent who knows he should face these fears but still retreats — he will also think that he acts permissibly or rightly in fleeing these situations. If he has the concept of cowardice at all, he will at least lack mastery of it or have a mistaken conception that leads him to consistently misjudge how he should respond to fearful situations. Or take the example of vices connected to racism, such as those identified by Lawrence Blum: racial ill will, racial inferiorizing, and racial disregard (Blum 2007). These vices may involve the habitual use of racist concepts, such as *NEGRO*. Mastering this concept as someone who uses it earnestly requires various beliefs about the stereotypical qualities of black people, and a belief that because they have such qualities and belong to this racialized group, they should be subject to discriminatory practices. The agent who earnestly and routinely uses the concept *NEGRO* will be driven to racist actions manifesting racial ill will, inferiorizing, or disregard, because the use of *NEGRO* presents black people as deserving such treatment.

The harmful behaviour driven by racist and similar thick ethical and social concepts is part of a broader range of the wrongful consequences of using such concepts. Consider, for instance, what Charles Mills says about the concept of a ‘savage’ as applied to indigenous people:

When Thomas Jefferson excoriates the ‘merciless Indian Savages’ in the Declaration of Independence... neither he nor his readers will experience any cognitive dissonance with the earlier claims about the equality of all ‘men’, since savages are not ‘men’ in the full sense... Even a cognizer with no antipathy or prejudice toward Native Americans will be cognitively disabled trying to establish truths about them insofar as such a category and its associated presuppositions will tend to force his conclusions in a certain direction... ‘Savages’ tend to do certain things and to be unable to do others; these go with the conceptual territory. Thus the term itself encourages if not quite logically determines particular conclusions... foundational concepts of racialized difference, and their ramifications in all sociopolitical spheres, preclude a veridical perception of nonwhites and serve as a categorical barrier against their equitable moral treatment. (Mills 2007, 27)

SAVAGE is plausibly a thick social concept in the sense described in the previous section. Mastering it requires the subject to have a series of beliefs about indigenous people that predicate various stereotypical descriptions — that they are prone to violence, simple-minded, and have a quaint reverence for nature. At the same time, the concept marks indigenous people as deserving discriminatory treatment on the basis of the ascription of these stereotypical qualities: that white people should be on their guard in case they pose a violent threat; that the colonial regime should attempt to ‘civilise’ them by educating them out of their traditional culture; that because they will not exploit the natural resources or space in their lands, developers may overlook their property rights.

Mills describes two interconnected problems arising from the use of the concept SAVAGE. The first is epistemic: thinkers who use this concept will find it difficult to arrive at true beliefs about indigenous people, because mastering the concept SAVAGE involves acquiring a variety of *false* beliefs about the ways in which these people behave. Just as applying a thick ethical concept provides the agent with *pro tanto* moral reasons to act, applying a thick social concept like SAVAGE provides the subject with putative epistemic reasons to believe that these stereotypical descriptions apply to particular indigenous people. Additionally, the concept encourages misinterpretations of the significance of various *facts* about indigenous people. For example, the fact that indigenous people are overrepresented in the Canadian prison system will, to someone who uses the concept SAVAGE, appear to be confirmation of the stereotype that indigenous people are violent — and not, for instance, a symptom of systematic discrimination. And, since there is no such thing as a ‘savage’ *per se*, it may be that the concept itself is empty: propositions containing the term may fail to refer, and thus be either systematically false or truth-invaluable.

The second problem is moral and political. Because the concept SAVAGE is embedded in racist social practices and institutions, the person who has mastered this

concept will be driven to act in problematic ways. Where we have the problematic beliefs that come along with mastery of the concept SAVAGE, our deliberations will draw upon these beliefs, making morally problematic discriminatory actions seem reasonable. The alarming incidence of violence towards indigenous people in the United States and Canada points to this effect.<sup>32</sup> At the social-political level, these beliefs will also serve as an impediment to engaging with indigenous people as equals and to seriously engaging with their needs and interests. Where one thinks of another group as simple and in need of civilizing, one will be predisposed to misconstrue and ignore the actual stated needs and desires of the members of that group in favour of the paternalistic image one already has. As Richard H. Pratt put it, schools should take the indigenous person, 'kill the Indian in him, and save the man' (Pratt 1892, 46). The conclusion of the Truth and Reconciliation Commission of Canada (2015) that the Canadian residential school system was a programme of 'cultural genocide' follows straightforwardly from the ambition that Pratt and others so bluntly stated.

However, a concept or conception need not issue in bad behaviour to be morally wrongful. Murdoch claims that the use of certain concepts can be morally wrongful even when confined to private thoughts and feelings. Despite the distaste that the mother-in-law feels towards her daughter-in-law, Murdoch tells us that '[M] behaves beautifully to the girl throughout, not allowing her real opinion to appear in any way. We might underline this aspect of the example by supposing that the young couple have emigrated or that D is now dead: the point being to ensure that whatever is in question as *happening* happens entirely in M's mind' (Murdoch 1970, 17). As

---

<sup>32</sup> It would probably be incorrect to suggest that every white person implicated in this violence does so because they think of indigenous people as 'savages'. But it is plausible that a racist *conception* of indigenous people persists, where many of the same beliefs and evaluations involved in SAVAGE continue to be held with regard to indigenous people. Depending on how we prefer to individuate concepts, this may indicate that the concept I have labeled 'SAVAGE' is indeed operative here, regardless of the words used by the actual thinkers involved.

mentioned above, Murdoch considers the change in M's private thoughts and feelings — e.g., from thinking of D as 'undignified' to thinking of her as 'spontaneous' — to be a moral improvement effected by a shift in the concepts M uses to think about D. But both the moral wrong of M's initial thoughts and feelings and the moral good of correcting them takes place without any practical upshot.

Similarly, Blum notes that racist vices can manifest entirely in one's private thoughts and feelings:

racial ill will or disregard also manifest themselves in certain feelings or emotions that do not necessarily prompt action. Delighting in the ill fortune of the racial other, anger or dismay at the racial other's successes, aversion to the presence of the racial other, glee when the racial other is humiliated, consternation that one's offspring or friend has befriended a member of a stigmatized race are or can be examples of such emotions... it would be appropriate to attribute vicious racial attitudes to someone who never actually engaged in racist actions but who nevertheless thought of another racial group as inferiors, or who wished them ill. (Blum 2007, 228)

So, merely thinking about indigenous people using the concept *SAVAGE* can manifest a racist vice, and thus be morally wrongful. Indeed, from a deontological perspective, it may be these inferiorizing thoughts that are the most morally egregious aspect of racism, for such thoughts disrespect the humanity of racialized people. As Ernesto Garcia writes, 'What seems most truly evil about racist action is not so much any particular acts of taunting, racial slurs, or displays of disrespect, but instead, we might feel, the pernicious underlying attitude: that the racist sincerely believes members of a different race somehow count as less than human' (Garcia 2002, 202–3). Again, with regard to *SAVAGE* and similar concepts, Mills writes: 'Whiteness is originally coextensive with full humanity, so that the nonwhite Other is grasped through a historic array of concepts whose common denominator is their subjects' location on a lower ontological and moral rung' (Mills 2007, 26). As discussed above, this dehumanization manifests in harmful actions, but it still wrongs racialized people even if it goes no further than the thoughts of a closeted racist.

Finally, the use of certain social concepts may restrict the kinds of social identities we can have and the lives we may live. Because concepts allow us to think about their objects in a particular way, without the required concepts it may be difficult or even impossible to classify certain social experiences, such as social identities that fall outside the norm. Trans people, for instance, often describe the moment they learned about the concept of being trans gender as a watershed event that made sense of their experiences in a way they had not previously been able to express. As Nikki Hayden, a 26 year old psychology student, told *The Guardian*: ‘I learned what trans meant through YouTube. I knew how I felt but I didn’t know there was a term for it. I was basically just trying to Google what I felt. A lightbulb went off in my head and I thought, this explains all the issues I’ve had as long as I can remember’ (Lyons 2016).<sup>33</sup> At the same time, our social concepts and conceptions are embedded in cultural practices, some of which may have a vested interest in maintaining certain social roles as they stand. For example, when trans activist Tara Hudson was on trial for head-butting a barman, she was initially jailed in a men’s prison. She later sued the British Ministry of Justice for the ‘intimidating, hostile, degrading, humiliating, and offensive’ treatment she experienced, and was shocked to find the government’s defence stated that ‘Hudson “is as a matter of biological fact a man” and... that the term “transwoman” has no legal significance’ (Townsend 2018). The absence of the concept TRANS WOMAN in British law thus serves to reinforce traditional conceptions of gender, in potentially dangerous ways.

More disturbingly, it is plausible that certain reactionaries have a thick conception of trans people that motivates the often violent actions they take to reinforce social norms that exclude the possibility of being trans. In Aceh, Indonesia, for example, police and locals arrested a group of trans women, stripped and beat

---

<sup>33</sup> I will come back to instances where a conceptual absence makes it difficult to interpret one’s own social experience in Chapter 2.



them, and forced them ‘to undergo an ad hoc form of “gender re-education”, which included wearing men’s clothing, physical exercises such as push-ups and sit-ups, and instruction on lowering their voices, until they were deemed suitably “macho”’ (Lamb 2018). Such violence and humiliation is part of an ongoing effort in parts of Indonesia (and similar efforts elsewhere) to enforce a certain traditional conception of gender roles, and is arguably motivated by a thick conception of trans gender people that marks them as deserving discriminatory treatment. Where such concepts are the norm, people may be forced — either by the absence of adequate ways of conceptualizing their own social identities or via coercion motivated by derogatory thick conceptions of those identities — into identities that have been socially constructed in ways that are against their interests.

Noxious concepts may also be invented out of whole cloth by reactionaries to enforce their preferred social order, to the detriment of marginalized groups. For example, the concept of a ‘race traitor’ was initially devised by Jim Crow era racists in the U.S. as a pejorative for white folks who supported civil rights legislation and activism. As Naomi Zack recounts, this concept not only picked out and evaluated a certain range of responses to civil rights legislation, it was also embedded in a commitment to racist institutions and to a substantive view of the social ontology of race:

This usage implied that loyalty to whites by other whites was a virtue comparable to patriotism and that membership in a hereditary caste could be betrayed by behavior disapproved of by other caste members... It is striking (as Martin Luther King emphasized) that the traitorous behavior was not unlawful according to federal or, in some cases, state law... Thus, the so-called white race traitor was someone who insisted on obeying the law when other members of his or her race insisted on disobeying the law. The treason at issue was, in short, a violation of a criminal code according to the criminals who believed they were above the law. (Zack 1999, 78)

After the civil rights movement, in a dark reflection of the anti-racist and feminist movements’ shift of emphasis from civic equality to social equality, the white supremacist conception of a ‘race traitor’ seems to have shifted from targeting

supporters of anti-racist legislation to targeting people whose social relations cross racial lines — especially people in mixed-race romantic relationships.<sup>34</sup> This deep connection to a racist worldview, Zack argues, makes the term ‘race traitor’ unlikely to be reclaimable by white anti-racist activists.<sup>35</sup> In trying to preserve and advance a racially hierarchical (or at least segregationist) social order, the concept of a ‘race traitor’ defines a range of social identities and ways of living that are explicitly opposed to egalitarian causes.

#### 1.4. Objections

To sum up the account so far: I have introduced and defended a working account of concepts, concept mastery, and conceptions; I have focused specifically on thick ethical and social concepts; and I have expounded various ways in which the use of certain concepts can be wrongful. Before moving on, there are three objections that are worth addressing. The first asks why we should concern ourselves with the moral appraisal of concepts at all, rather than the moral appraisal of attitudes. The second asks whether my account so far might generate impossible demands, or whether it may turn out to be self-effacing. The third worries that whatever moral demands my account raises may open the way to thought policing.

##### 1.4.1. *Why Concepts?*

Several times in the discussion above, I referred to the morally or epistemically problematic attitudes that one must have in order to have mastered a concept, or that one is likely to acquire by using that concept. We might think that this pattern shows where the real problem lies in these cases. That is to say, one might object that the morally significant aspect of these cases is not the concepts that we use, but the

---

<sup>34</sup> See, for example, this discussion on the white supremacist forum ‘Stormfront’: <<https://www.stormfront.org/forum/t1198329/>>, accessed 27 Mar 2018.

<sup>35</sup> See Ignatiev and Garvey (1996) for such an attempt.

attitudes we acquire. It is not our tools that are the problem; rather, it is the ways in which they are used.

But in many cases where the kinds of concepts I have been describing are in operation, the only locus of moral impropriety will be the concept that is used, and *not* the attitude in which it figures. The concept SAVAGE, for example, may be used to construct many beliefs that are not obviously problematic *apart from* the use of that concept. When the subject sees an indigenous person and acquires the belief that ‘there is a savage over there’, it is hard to see what could be morally objectionable apart from the evaluative content implied by the use of the concept SAVAGE. If the concept turns out to be non-empty — that is to say, despite identifying a social category in a morally problematic way, the concept still succeeds in fixing a reference — then it is not entirely clear that there anything *epistemically* problematic with this belief, either. Theorizing that the concept itself is morally problematic accounts for the moral wrongfulness both of attitudes with no denigrating content apart from the offending concept (‘there is a savage’), as well as the more central cases of attitudes that make the denigration explicit (‘savages are primitive’). Mastery of this concept may consist in having various problematic attitudes, but mere use of the concept may be sufficient to produce these moral wrongs.

#### 1.4.2. *Impossible Demands?*

If concepts can be morally wrong, one might read this as suggesting that we have a moral obligation to *eliminate* those concepts. But, an objector might say, this suggests that my account demands that we not possess the offending concepts at all. And this, they continue, is to make a moral demand that is either impossible to meet — for it is not clear how one could remove a concept from one’s mind — or else makes my account self-effacing — for if we are required not to have certain concepts, we could not understand and be motivated by the prescription against them.

First of all, though this is *not* the view I take, it is not entirely clear that these would be problems for my account. I do not have an answer as to how one could remove a concept from one's repertoire, but we should not assume that this is impossible. Even if it is impossible to scrub a concept from one's repertoire, we might interpret the moral demand to eliminate those concepts not as demanding us to do this impossible task (violating the principle that 'ought' implies 'can'), but rather as demanding that we bring up future generations such that they never acquire these concepts in the first place. As to self-effacement, there are two potential responses. On the one hand, it is not obvious that a moral theory that prescribes against being motivated by the theory's own conclusions is problematic — many consequentialists, for example, happily admit that their theories are self-effacing.<sup>36</sup> On the other hand, if the demand to eliminate certain concepts concerns not our own conceptual repertoires but those of future generations, then the theory is not self-effacing after all: we may be motivated by its prescriptions in ensuring that our offspring do not acquire these problematic concepts, and they will passively meet the theory's demands without needing to be motivated by it at all.

But we can set these responses aside, because I do not think that my account makes the problematic moral demands the objector points to here. I am committed to nothing so strong as the view that we must eliminate wrongful concepts. Indeed, I think that would be a bad thing to do, all things considered. It is important that we retain some level of comprehension of these concepts in order to understand the thoughts and actions of people who still have full mastery of the concept and use it in earnest, as well as to interpret the thoughts and actions of people in the past. If a concept is morally wrong, the right thing to do would *not* be to remove it from our minds, but to

---

<sup>36</sup> For the classic argument against self-effacing moral theory, see Stocker (1976). For some consequentialist admissions of self-effacement, see Hare (1988, 289–90), Mill (1979, 17–19), Parfit (1984, 24, 43–45), Railton (1984), and Sidgwick (1981, 405).

retire it from use except for purposes of interpretation or, possibly, irony. This weaker demand essentially says that when a concept is morally problematic in the ways I have been considering, we should adopt a different *conception* in its place: retain our understanding of how that concept was used and its (now-)historical relation to various social practices, but abandon the beliefs that came along with mastery of that concept. For example, with regard to SAVAGE, my account does not demand that we remove this concept from our minds — that we should make it impossible to even think of someone as a ‘savage’. Rather, my account holds that we should not have the problematic *beliefs* that come with mastery of this concept. We may retain a diminished *conception* of a ‘savage’ for historical interpretation, mastery of which would involve not beliefs about indigenous people but rather beliefs about the beliefs of white people who thought of indigenous people as ‘savages’. This position corresponds to the situation Williams thinks we may arrive at after critical reflection takes away our conviction that a concept and its associated social practices are part of a good way to live: ‘if [we] think about [our] earlier beliefs, [we] will now see them as the observer saw them, as knowledge [we] do not share’, because we no longer have the same conception that we did before (Williams 1985, 167).

#### 1.4.3. *Thought Police?*

The notion that our concepts could be subject to moral evaluation may raise familiar dystopian images: are morally bad thoughts *thoughtcrimes*, which we must discipline ourselves into avoiding by practicing *crimestop* and acquiring a purified language such as *Newspeak*, all of which may be enforced by the *thought police*?<sup>37</sup> At least, so an objector might say, in an admittedly polemical tone. As I understand it, there are three more charitable ways to state the objection.

---

<sup>37</sup> This scenario is, of course, from George Orwell’s *Nineteen Eighty-Four* (1949).

The first form of this objection takes issue with the notion that moral evaluation should go as far as our private thoughts and feelings. The idea here is that these aspects of our lives are not subject to moral evaluation, perhaps because they are not directly observable by others, or because they do not in themselves constitute harmful behaviour. So, the objector concludes, my account is engaged in a kind of moral overreach. But I simply do not see the force of this objection; it strikes me as an attempt to selectively avoid accountability for an area of moral life.<sup>38</sup> Our 'inner life' may not be directly observable by others, but they may infer a great deal about our thoughts by listening to what we say and observing what we actually do. Moreover, the notion that our actions must be observable for others to morally evaluate in order for those actions to be morally significant is absurd: would Robinson Crusoe torturing sea creatures out of boredom be any less reprehensible if no one but him ever knew? And as I discussed above, while it is conceivable for problematic concepts to have no effect beyond the subject's private thoughts, in many cases these concepts *will* have an effect on the subject's behaviour.

The second form of this objection concerns the possibility that moral injunctions against using certain concepts may devolve into violations of rights to freedom of thought and expression. But it is worth reminding ourselves that these are political rights guaranteed against the *state* — we have a right that the *government* not try to interfere with what we think and say. And enforcing morality is not something we delegate wholesale to the state. While many laws have a basis in moral norms, it would be perverse for the state to mediate every aspect of our moral lives. My view does not imply that books should be censored and 'thoughtcrimes' should be prosecuted, any more than the immorality of theft implies that books about thievery should be banned or that the state should arrest me for stealing a potato chip from my

---

<sup>38</sup> Indeed, Murdoch (1970) seems to make just this accusation of expressivist and existentialist moral theories.

partner's lunch. It may, however, have implications for other things the state may do to encourage moral behaviour: for example, in an effort to reduce the spread of problematic racial concepts, perhaps schoolteachers should be careful to place colonial-era books in context and explain that the tacitly and overtly racist views expressed in them are wrong. If the worry is about such softer methods the state or influential social groups might take to influence our thinking, I take it that this worry only has bite given the possibility that we may be mistaken about the wrongfulness of thinking in one way or another. But this is easily remedied. It is a basic principle of modern democratic societies committed to freedom of thought and expression that we must leave some freedom to try out different ways of thinking and living. As Williams points out, it would be wrong to try and stop this:

We should not try to seal determinate values into future society... We should try to leave resources for an adequate life and, as means to that and as part of it, we shall try to transmit what we take to be our knowledge [and concepts]. We cannot consistently leave out the reflective consciousness itself and practices of free inquiry needed to sustain it and to make use of it. (Williams 1985, 173)

We can do our best to leave future generations what we take to be our best practices and concepts, but they must take this up for themselves.

The third form of this objection concerns the possibility of *moralism*. The worry here is that although we can accept that certain ways of thinking are morally problematic, nevertheless we should refrain from criticizing people for these errors. The idea here is perhaps that blame or criticism would be too harsh a response to conceptual wrongs. There are three ways of responding to this worry. First, note that moralism is a danger that appears in all areas of moral life. Our ways of holding one another accountable for moral failings always have the potential to be exploited in self-serving, abusive, or authoritarian ways.<sup>39</sup> Merely bringing the ethics of our conceptual practices into view does not imply that this degeneration will occur. Second, it would

---

<sup>39</sup> Cf. Fricker (2018).

be strange to bar us from ever criticizing people's concepts on moral grounds. Moral criticism is an important means by which we can effect moral improvement, and there is no reason to think this is any less true of our concepts. This is particularly the case given that, as I argue in Chapter 2, structural injustices lead to widespread *ignorance* of various concepts that are important for understanding the experiences of marginalized groups. Third, the injunction not to criticize people for the use of wrongful concepts seems to assume a substantive view regarding our *moral responsibility* for the use of said concepts. If it were to turn out that we are not morally responsible for using these concepts, then blame and perhaps criticism for using them may not be appropriate. But we cannot simply assume this. Indeed, in Chapter 3, I will argue that on a preferred view of moral responsibility, we *are* responsible for our use of problematic concepts. Blame and criticism would thus be appropriate after all.

### 1.5. Summary

In this chapter, I have argued that some of our concepts may be morally problematic in various ways. After outlining a working account of concepts that is compatible with three major theories, I moved specifically to consider thick concepts — those with both evaluative and descriptive aspects. These concepts may be problematic to use because they partially constitute vices, issue in harmful action, construct wrongful (potentially dehumanizing) thoughts, or construct social reality such that certain social groups are constructed in a way contrary to their interests. I also considered objections to the effect that we should concentrate on attitudes rather than concepts, that moral evaluation does not concern 'inner' mental life, and that the possibility of making such evaluations opens the way to an oppressive régime of thought policing. In the next chapter, ahead of moving to the question of our moral responsibility for our concepts, I discuss the moral and epistemic problems that arise because of *conceptual ignorance* — a *lack* of certain concepts, conceptions, or concept mastery.



## 2. Conceptual Ignorance and Epistemic Injustice

In the previous chapter, I outlined some of the moral, epistemic, and social problems that arise from the use of certain concepts, such as SAVAGE or RACE TRAITOR. In this chapter, I introduce a complementary range of moral, epistemic, and social problems that arise from *not* using certain concepts. In the first section, I define two ways in which the subject might fail to properly employ the relevant concepts: conceptual ignorance and conceptual non-competence. In the second section, I describe a structural source of conceptual ignorance and non-competence that generates moral and epistemic harms: hermeneutical injustice. In response to recent scholarship on hermeneutical injustice, I propose an account of conceptual resources that are specific to various social groups, and the kinds of hermeneutical injustice that can result when different social groups lack the required concepts. In the third section, I shift back from social-epistemic structures to individual actions, considering the notion of wilful hermeneutical ignorance. I argue that this category cannot account for many intuitively culpable instances of conceptual ignorance or non-competence, setting up the next chapter's discussion of moral responsibility for our conceptual incapacities.

### 2.1. Conceptual Ignorance and Conceptual Non-competence

Public discussion of sexual harassment is at an all-time high. Alongside prominent accusations against men in business, entertainment, and politics, there is a fairly substantive public debate taking place regarding which sorts of behaviour count as harassment. But it is worth noting that the very concept, SEXUAL HARASSMENT, is relatively young, the term having been invented by feminist activists sometime in the 1970s.<sup>40</sup> Similar behaviour had been the focus of public discussion before: in the late nineteenth and early twentieth centuries, for instance, American newspapers decried

---

<sup>40</sup> For one account of the mid-1970s origin of the term 'sexual harassment', see activist Susan Brownmiller's memoir (Brownmiller 1990).

both the sexually harassing behaviour of men in public places (which at the time was called ‘mashing’) and the sometimes violent ways women defended themselves — on more than one occasion, it was reported that women were jabbing mashers with their long hat pins.<sup>41</sup> But between the Second World War and consciousness-raising in the 1970s and 80s, the issue seems to have receded from public consciousness as traditional gender roles were reasserted in popular culture. Even after feminist activists and academics began using the term ‘sexual harassment’ and adopting policies to prevent and punish such behaviour, the watershed moment for public awareness did not come until Anita Hill’s testimony, in 1991, against the nomination of Supreme Court Justice Clarence Thomas, wherein she accused Thomas of making remarks to her that constituted sexual harassment.<sup>42</sup> Since the very idea of sexual harassment is a 20th-century (or, at the earliest, 19th-century) conceptual innovation, and furthermore, since widespread awareness and use of this concept seems to have come about only in the last thirty years or so, it follows that many people did not have or did not know how to properly apply the concept of sexual harassment until recently.

Recall the account of concepts that I introduced in the previous chapter. I introduced a distinction between concepts (the constituents of thoughts), concept mastery (the range of beliefs, knowledge, or other cognitive requirements for competently applying a concept), and conceptions (different arrays of beliefs and knowledge that may accrue to the same concept in different contexts). To possess a concept of X, the subject must be able to identify things as X’s and to acquire attitudes structured by that concept. If the subject *does not possess* some concept, we may say that she is *conceptually ignorant*. A subject who is ignorant of the concept SQUARE, for instance, would not be able to identify objects as squares or to entertain a belief that

---

<sup>41</sup> For more on ‘mashers’, see Kerry Segrave’s history of sexual harassment in the U.S. from 1880–1930 (Segrave 2014).

<sup>42</sup> For Hill’s account of this event, see Hill (1997). For Thomas’s account, see Thomas (2007).

some object is a square. She would not be able to think about square objects *as* squares, though she could think about them as something else.<sup>43</sup>

If the subject instead *lacks mastery* of a concept — if she lacks the appropriate range of knowledge and beliefs about the object of that concept, knowledge of how to apply the concept, and knowledge of the inferences that using that concept entitles one to make — we may say that she is *conceptually incompetent*, or, less pejoratively, *conceptually non-competent*. Mastery of the concept SQUARE may involve knowing that when one identifies an object as a square, one may conclude, for example, that it has four interior right angles. The subject may possess a concept while remaining non-competent with respect to that concept: children, for example, learn to identify squares long before they learn that all squares have four interior right angles. The subject might also be conceptually non-competent because she has a mistaken *conception*, that is to say, false beliefs regarding the upshot of using that concept. A child might mistakenly believe that squares are not rectangles, for example, and thus fail to know that she is entitled to conclude of any square that it is a rectangle.

Given how fundamental basic shapes are to our experience of the world, it seems unlikely that many people are ignorant or non-competent with regard to SQUARE. But when it comes to more complex and contentious entities, such as social kinds, conceptual ignorance and non-competence become a live issue. In the case of the concept of sexual harassment, for example, if one does not possess this concept (or something similar, like the idea of ‘mashing’), then one may be at a loss to identify the behaviour it picks out as a distinctive class of wrongful action. One may even be driven to identify this behaviour as something else — for example, as ‘flirting’ — that completely obscures its problematic nature. And even if one has the concept of sexual

---

<sup>43</sup> If all (or some) concepts are innate, then it is not strictly correct to speak of not possessing (those) concepts. In that case, it may be satisfactory to speak instead of an inability to use that concept, or of a lack of concept mastery (see below). See Fodor (1975, 1981) for a defence of conceptual nativism.

harassment, if one lacks mastery of this concept or has a mistaken conception of it, then one may fail to identify certain instances of sexual harassment (thinking, perhaps, that inappropriate touching counts but inappropriate remarks do not), or perhaps even think that this kind of behaviour is not actually wrongful (for instance, a misogynist who thinks he is entitled to women's bodies may understand how the term 'sexual harassment' is used but at the same time believe that it is inappropriately moralizing acceptable behaviour). Conceptual ignorance and non-competence are thus impediments to knowledge of the objects the relevant concepts pick out, and can therefore obscure social and moral problems.

The causes of conceptual ignorance or conceptual non-competence, as of ignorance more generally, are varied and may be more or less benign.<sup>44</sup> On the one hand, conceptual ignorance and non-competence may be the result of circumstantial factors, which could be instances of circumstantial bad luck or the result of patterns of oppression. A woman living in Ancient Greece would be ignorant of the concept of sexual harassment because of her historical context: the very idea of sexual harassment has not yet been developed there. It is a matter of bad circumstantial luck that this concept is simply unavailable to her. By contrast, a woman living in a conservative community in the 1990s may be conceptually non-competent with respect to SEXUAL HARASSMENT (suppose she is aware of the idea but believes it to be a fabrication of progressive hysteria) because of an oppressive social environment, where women's experiences are routinely ignored or silenced.

On the other hand, ignorance and non-competence may be the result of personal factors, which may be matters of constitutive bad luck or of personal moral-

---

<sup>44</sup> See Le Morvan and Peels (2016) for two candidate analyses of ignorance as lack of knowledge or as lack of true belief. They consider notions of knowledge and ignorance with regard to facts (factual), objects and people (objectual), and skills (procedural), but not of concepts. On my account, conceptual ignorance is perhaps a form of factual or objectual ignorance, while conceptual non-competence may be a form of procedural ignorance.

epistemic fault. A psychopath who is constitutively unable to grasp any moral concepts may be conceptually non-competent with respect to SEXUAL HARASSMENT, because he is unable to acquire the range of moral beliefs that are required. But that conceptual non-competence is not the psychopath's fault; it is a matter of constitutive luck that he is unable to master moral concepts. By contrast, a man who faces no such constitutive barriers to comprehension, and has heard women describe sexual harassment as a moral wrong, yet continues to lack mastery of the concept SEXUAL HARASSMENT (perhaps failing to think of it as morally wrongful behaviour), seems morally and epistemically culpable for his conceptual non-competence. In this case, mastering the concept is within his grasp, but he has not done what he (morally and epistemically) ought to do.

I return to the question of culpability for conceptual ignorance or non-competence in the third section of the present chapter, and discuss this in more detail in Chapter 3. I discuss issues of bad moral-epistemic luck in Chapter 4. The remainder of the present chapter presents some causes of conceptual ignorance and non-competence that are effects of structures of oppression or wilful disregard. In the process, the ways in which conceptual ignorance and non-competence can be morally and epistemically wrongful will be apparent.

## 2.2. Structures of Conceptual Ignorance: Hermeneutical Injustice

People in certain professions and social locations, such as journalists, politicians, academics, entertainers, and legal professionals, have greater cultural influence than others. Wilfully or not, these people's interests and experiences tend to set the agenda for culturally significant conversations. But people from certain social groups are systematically excluded from or underrepresented in these influential locations: the listed professions are relatively élite and historically limited to members of more privileged social groups. As a result, the social experiences of these marginalized groups may be excluded from widespread cultural awareness. Miranda Fricker refers to this situation as *hermeneutical marginalization* (Fricker 2007, 153 ff.). This structural

disadvantage puts the members of hermeneutically marginalized groups at an unfairly heightened risk that the collective hermeneutical resource — the shared stock of interpretive tools — will lack the appropriate concepts and conceptions for making sense of their distinctive and important social experiences.<sup>45</sup>

These collective hermeneutical gaps create a situation where many if not most subjects will be conceptually ignorant or non-competent with regard to the concepts and conceptions that are needed to make proper sense of the distinctive social experiences of hermeneutically marginalized subjects. Where such ignorance or non-competence impedes the understanding of these experiences, Fricker argues that the marginalized subject experiences a *hermeneutical injustice*: ‘the injustice of having some significant area of one’s social experience obscured from collective understanding owing to persistent and wide-ranging hermeneutical marginalization’ (Fricker 2007, 154). The primary harm of hermeneutical injustice is epistemic: ignorance or non-competence with respect to the required concepts and conceptions prevents one from acquiring and sharing knowledge of those social experiences. There are also a variety of secondary harms. Some of these harms are practical: without being able intelligibly to articulate one’s important social experiences, one will find it difficult to respond to them when they arise — potentially leading to moral harm. Other secondary harms are epistemic: persistent experiences of hermeneutical injustice can undermine one’s epistemic confidence or cause one to lose one’s conviction in knowledge that one has, downgrading the epistemic status of those attitudes to a suspension of judgement or even disbelief. Another kind of secondary harm is social-ontological: Fricker describes a case from Edmund White’s semi-autobiographical novel, *A Boy’s Own Story*, where the protagonist experiences such persistent

---

<sup>45</sup> These interpretive tools also include linguistic and expressive resources, such as words, phrases, and dialects. In addition, Charlie Crerar (2016) argues that hermeneutical resources should include ‘expressively free environments’.

interpretive difficulties in coming to terms with his identity as a gay man that he internalizes some of the noxious stereotypes about homosexuality (White 1983). This identity-constructive harm of hermeneutical injustice, Fricker argues, may prevent the subjects of hermeneutical injustice from 'becoming who they are' (Fricker 2007, 168).<sup>46</sup>

However, it is unclear whether members of marginalized groups who have successfully produced their own concepts and conceptions in spite of the collective gap also suffer hermeneutical injustice. Consider, for example, the multitude of new concepts and terms that have been produced in LGBTQ+ communities, such as the idea of being genderless or agender. Members of these communities may well have mastered the concept AGENDER and thus understand a person's experience of genderlessness perfectly well. But many who are differently situated continue to struggle to make sense of this non-binary gender identity. As agender activist Tyler Ford writes in an autobiographical essay for *The Guardian*:

I [came] across the concept of non-binary gender identities while reading blogs written by trans people. At this point, gender fluidity and gender neutrality was not being discussed in the media as it is today by celebrities such as Miley Cyrus and Shamir Bailey... Back then, I had never seen the topic addressed publicly, or by anyone in my life... I have been out as an agender, or genderless, person for about a year now. To me, this simply means having the freedom to exist as a person without being confined by the limits of the western gender binary. I wear what I want to wear, and do what I want to do, because it is absurd to limit myself to certain activities, behaviours or expressions based on gender. People don't know what to make of me when they see me, because they feel my features contradict one another. They see no room for the curve of my hips to coexist with my facial hair; they desperately want me to be someone they can easily categorise... Strangers are often desperate to figure out what genitalia I have, in the hope that my body holds the key to some great secret and unavoidable truth about myself and my gender. It doesn't. My words hold my truth. My body is simply the vehicle that gives me the opportunity to express myself. (Ford 2015)

---

<sup>46</sup> The remainder of this section is a lightly edited version of the central sections of Goetze (2018). The majority of changes are to bring out this chapter's focus on conceptual ignorance and non-competence, but I have also made a few minor changes and additions for clarity. The reproduced text is used with permission from the publisher.

We can call cases like Ford's, where the members of a hermeneutically marginalized group generate their own concepts and conceptions, instances of *hermeneutical dissent*. There are many other examples from feminist, queer, race, and disability theory, where hermeneutical dissent has overcome a collective interpretive failure.<sup>47</sup> But are Ford and others like them subject to hermeneutical injustice? According to several critics of Fricker's view, the answer would seem to be negative. Drawing on examples of hermeneutical dissent, José Medina, Rebecca Mason, and Kristie Dotson each argue that Fricker's account of hermeneutical injustice either fails to capture or rules out the possibility of hermeneutical dissent. If the former is true, then this is a serious oversight; if the latter is true, then Fricker's account runs against some of the core commitments of feminist and other liberatory theory. Yet, hermeneutical dissenters like Ford seem to suffer some of the same harms as the subjects of hermeneutical injustice: they are prejudicially excluded from sharing their knowledge of their own experiences due to their hermeneutical marginalization. To get a clear grasp of both the structural causes of conceptual ignorance and non-competence, and the moral and epistemic harms thereof, we need to get clear on the nature of hermeneutical injustice and hermeneutical dissent.

In this section, I argue that we can account for cases of hermeneutical dissent as instances of hermeneutical injustice, provided we make three clarifications. First, to make our account of hermeneutical injustice compatible with hermeneutical dissent in the first place, I clear up our understanding of the collective hermeneutical resource, arguing that it consists in the concepts and conceptions that are shared by all. Second, to show how hermeneutical dissenters can still be subject to a hermeneutical injustice, I dissolve some ambiguities surrounding the primary harm of hermeneutical injustice, arguing that the primary harm is that the subject's experience lacks intelligibility,

---

<sup>47</sup> See, for example, the feminist standpoint tradition, summarized by Anderson (E. Anderson 2015a, sec. 2).



*either cognitively or communicatively*. Third, I provide a more nuanced account of hermeneutical resources that makes it easier to track the extent of hermeneutical gaps relative to various social groups and individuals. This fine-grained picture of the hermeneutical economy allows us to distinguish six species of hermeneutical injustice.

### 2.2.1. *Collective Resources and Dissenting Communities*

According to Medina, Mason, and Dotson, the way Fricker presents the collective hermeneutical resource creates problems in accounting for hermeneutical dissent. Medina claims that ‘Fricker’s expression “the collective hermeneutical resource” strongly suggests that we can pool all the hermeneutical resources available to all groups’ (Medina 2013, 103). According to Medina, Fricker presents the collective hermeneutical resource as *cumulative*: it refers to *all* of the concepts and conceptions in circulation across *all* discursive communities. The interpretation presented by Mason and Dotson is slightly different. Mason claims that Fricker presents the concepts and conceptions used by members of socially powerful groups as the *only* hermeneutical resource available to marginalized subjects, and ‘thus glosses over important distinctions — in particular, distinctions between dominant and non-dominant hermeneutical resources’ (R. Mason 2011, 300). Dotson makes much the same claim: ‘Fricker seems to assume that there is but one set of collective hermeneutical resources that we are all equally dependent upon’ (Dotson 2012). According to Mason and Dotson, Fricker presents the collective hermeneutical resource as *exhaustive*: it is the only resource available. We can illustrate these interpretations using Venn diagrams (Figure 1): the shaded region is the collective hermeneutical resource and each circle represents some collection of concepts and conceptions. On Medina’s interpretation (a), the collective resource is the *union* of various community-specific hermeneutical resources; on Mason and Dotson’s interpretation (b), the collective resource is the *only* hermeneutical resource.

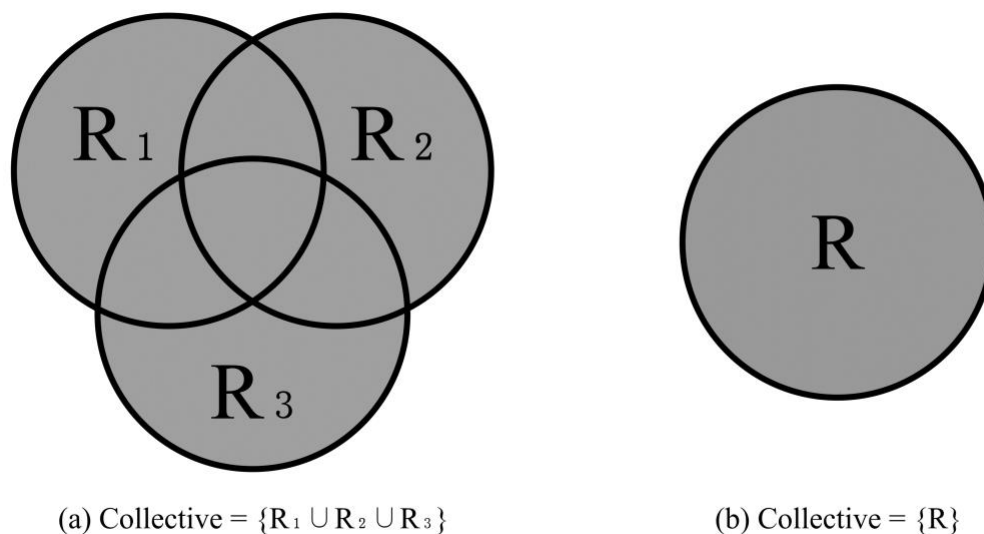


Figure 1. (a) Medina's interpretation. (b) Mason's and Dotson's interpretation.

Because a gap in the collective hermeneutical resource that has been produced by hermeneutical marginalization is the background condition for hermeneutical injustice, these interpretations necessitate that a hermeneutical injustice can occur *only* in cases where no interpretive tools that could make that experience intelligible are available *to anyone*. Thus, Medina finds it 'problematic that Fricker operates with the working assumption that when there is a hermeneutical gap, a range of experiences will be rendered unintelligible *for everybody*' (Medina 2013, 101, his emphasis); Mason claims that 'when a hermeneutical lacuna exists with respect to the experiences of a marginalized group, *everyone* fails to understand' (R. Mason 2011, 303, my emphasis); and Dotson makes the same point: 'for Fricker, the hermeneutical lacuna in hermeneutical injustice renders some experiences difficult to conceptualize *for the marginalized and the perceiver alike*' (Dotson 2012, my emphasis).

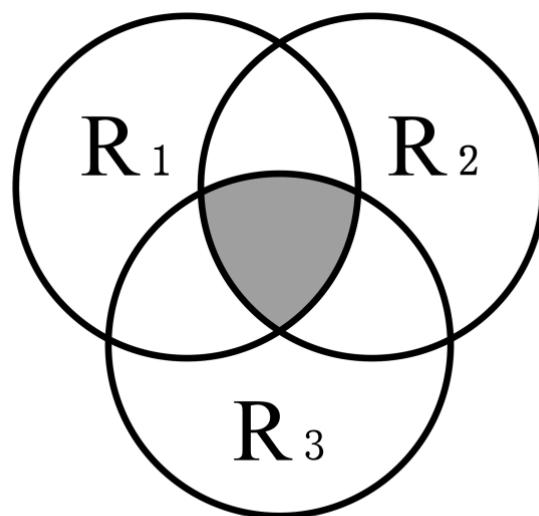
But if ignorance and non-competence with regard to the required concepts must be universal to generate a hermeneutical injustice, the presence of hermeneutical dissent *rules out* hermeneutical injustice. On Medina's reading, the collective resource extends across all the interpretive tools in circulation, regardless of who actually has them. Subjects in hermeneutically dissenting communities therefore avoid

hermeneutical injustice because they have produced the interpretive tools required to understand their experiences, despite the conceptual ignorance and non-competence of subjects in dominant social groups. But this gap in dominant hermeneutical resources leads to an interpretive breakdown when marginalized subjects attempt to explain their experiences to dominant subjects. Medina claims that the dominant subjects suffer an epistemic harm, in that they cannot obtain knowledge about the dissenters' experiences, yet, 'interestingly and crucially, the hermeneutical harms are *wrongful for others*, not for those upon whom the epistemic harms are directly inflicted' (Medina 2013, 107). This phenomenon, Medina concludes, is not addressed on Fricker's account of hermeneutical injustice. On Mason and Dotson's interpretation, however, hermeneutical dissent cannot even be accounted for. If the collective hermeneutical resource is the *only* conceptual resource available, then for any given concept, either everyone has it or no one does. Therefore, hermeneutical dissent is impossible — but since we can point to clear cases of hermeneutical dissent, it seems there must be something wrong in Fricker's account.

In a recent piece, however, Fricker has resisted these interpretations, claiming that 'a commitment to the existence of localised interpretive practices that may perfectly capture a given range of experiences but whose meanings are not sufficiently shared across wider social space is already present at the heart of the original account of hermeneutical injustice' (Fricker 2016, 167). That is to say, Fricker claims that her original account can already accommodate hermeneutical dissent. According to her, the collective hermeneutical resource is neither cumulative nor exhaustive: '[the collective hermeneutical resource] contains only meanings that just about anyone can draw upon and expect those meanings to be understood across social space by just about anyone else. [It] contains those concepts and conceptualisations that are held *in common*' (163, her emphasis). The hermeneutical gap is in the common resource that *everyone* has access to — and the existence of this gap does not at all imply that *no one* must have

access to the required tools. Again, we can illustrate with a Venn diagram (Figure 2).

According to Fricker, the collective hermeneutical resource is the *intersection* of various community-specific resources.



$$\text{Collective} = \{R_1 \cap R_2 \cap R_3\}$$

Figure 2. Fricker's model.

It is easy to place hermeneutical dissent on Fricker's intersectional account of the collective hermeneutical resource. The existence of a gap in the stock of concepts and conceptions shared by all says nothing about the interpretive tools that may be available among the members of any *particular* group. That group's hermeneutical dissent may successfully produce the concepts and conceptions that are needed to make intelligible sense of their experiences despite the gap in the collective hermeneutical resource. In fact, we can see hermeneutical dissent in action in some of Fricker's own examples from her earlier work. Consider Wendy Sanford's experience of postpartum depression as recounted by Susan Brownmiller (1990). In 1969, Sanford was an upper-class Republican in the U.S. struggling with depression after giving birth to her son, and she and her husband had been blaming her for these difficulties. A friend convinced Sanford to come along to a feminist consciousness-raising seminar, where she learned about postpartum depression, conceived as a medical condition and

not a personal deficiency. Sanford came away with an understanding of her experience that made better sense of her predicament and helped her to realize that the way she was feeling was not her fault. Fricker (2007, 148–49) describes Sanford as suffering a hermeneutical injustice: the collective hermeneutical resource did not provide the appropriate interpretive tools for Sanford to properly understand her experience, because women were hermeneutically marginalized. And yet, feminist activists *had* produced the needed interpretive tools, despite the collective gap and despite their hermeneutical marginalization *qua* women. By participating in the consciousness-raising seminar, Sanford was able to overcome her conceptual non-competence and acquire the needed understanding of postpartum depression. Clearly, Fricker intends her account to have room for hermeneutical dissent.

This defeats the version the worry raised by Mason and Dotson that hermeneutical injustice cannot accommodate hermeneutical dissent — but it does not yet fully answer Medina’s worry about cases where hermeneutical dissenters are unable to make themselves understood to dominant subjects. As was the case for agender people in the Ford case, it is unclear whether the feminist activists in the Sanford case suffer a hermeneutical injustice. That our account has space for hermeneutical dissent does not yet show that hermeneutical dissenters are subject to hermeneutical injustice.

### *2.2.2. The Harms of Hermeneutical Injustice*

In order to show that the interpretive breakdown between hermeneutical dissenters and ignorant subjects constitutes a hermeneutical injustice, we must examine the harms of hermeneutical injustice. We have already seen that a gap in the collective hermeneutical resource owing to the subject’s hermeneutical marginalization does some kind of harm to the paradigmatic subjects of both hermeneutical injustice and hermeneutical dissent. If the harms in both cases turn out to be the same, we ought to identify hermeneutical dissenters as suffering hermeneutical injustice as well.

But it may seem that the harms suffered in each case are distinct. On the one hand, some instances produce *communicative* harms. In these cases, the subject is not prevented from acquiring knowledge of her experience, but she is prevented from sharing that knowledge because others find her testimony unintelligible. On the other hand, the interpretations presented by Medina, Mason, and Dotson all concentrate on cases that produce *cognitive* harms. In these cases, the subject herself is unable to attain knowledge of her experience in the first place. Both the communicative harm and the cognitive harm undermine the subject's capacity as a knower — the former by impairing her ability to share knowledge, the latter by preventing her from acquiring knowledge. In order to understand whether hermeneutical injustice afflicts hermeneutically dissenting subjects, we must clear up this ambiguity.<sup>48</sup>

On the one hand, Fricker sometimes claims that the primary harm is communicative: 'the primary harm of hermeneutical injustice consists in a situated hermeneutical inequality: the concrete situation is such that the subject is rendered unable to *make communicatively intelligible* something which it is particularly in his or her interests to be able to render intelligible', which amounts to 'prejudicial exclusion from the *spread of knowledge*' (Fricker 2007, 162, my emphasis). We can see this most clearly in another of Fricker's examples, which she takes from Ian McEwan's novel *Enduring Love* (McEwan 1998). The novel's protagonist, Joe Rose, is stalked and harassed by Jed Parry, a religious zealot deluded by an erotomaniacal misapprehension that Joe is in love with him. At key moments in the story, Joe tries to communicate his experience of Jed's behaviour to other people — first to his partner Clarissa, then to the police — but each time he is unable to make his experience intelligible to his

---

<sup>48</sup> Rebecca Mason convinced me of this in her helpful commentary on an earlier version of this material presented at the 2016 Congress of the Canadian Philosophical Association.

interlocutors.<sup>49</sup> Importantly, Joe does not suffer a cognitive harm at all: he *knows* that there is something very wrong with and potentially dangerous about Jed. Nevertheless, Joe, through no fault of his own, is unable to explain Jed's menace in a way that others find intelligible. Joe is harmed in his capacity as a knower because he is prevented from sharing his knowledge.

On the other hand, Fricker sometimes claims that the primary harm is cognitive. One example is the case of sexual harassment alluded to earlier. Brownmiller (1990) describes the experience of Carmita Wood, a university employee who was sexually harassed by a male professor. Before Wood and other feminist activists came together to discuss these experiences, inventing the term 'sexual harassment' in the process, a collective hermeneutical gap prevented Wood or anyone else from attaining knowledge of this experience. As Fricker puts it: 'a patch of her social experience which it was very much in her interests to understand was not collectively understood, and so *remained barely intelligible, even to her*' (Fricker 2007, 162, my emphasis).<sup>50</sup> Another example is the Sanford case described earlier. Fricker describes the moment Sanford realized that she was suffering from postpartum depression as a 'revelation concerning an experience of female depression, previously *ill-understood by the subject herself, because collectively ill-understood*' (149, my emphasis). Wood and Sanford both were

---

<sup>49</sup> Unlike systematic cases where an entire social group is hermeneutically marginalized, tracking them across multiple aspects of their lives, Joe's marginalization is of an incidental type, affecting only him, in a very limited aspect of his epistemic life.

<sup>50</sup> It remains slightly obscure exactly what Fricker means when she says that an experience is 'barely intelligible'. It cannot be the case that the subject's experience is so radically unintelligible that it cannot figure in her thought at all; for then it would seem impossible that she could engage in hermeneutical dissent or any other interpretive activity regarding that experience. In some cases the subject's nascent understanding may amount to some form of epistemic achievement that remains less coherent than propositional belief, perhaps what Alexis Shotwell (2011) calls 'implicit understanding'. In other cases, such as Sanford's, the subject may instead have a perfectly intelligible yet false belief about her experience. Regardless, the subject is prevented from acquiring *knowledge* of her experience.

harmed in their capacity as knowers because they were prevented from acquiring knowledge of their experiences.

If we try to interpret either the communicative or the cognitive harm as the primary harm of hermeneutical injustice, we encounter problems accounting for all of the central cases *and* hermeneutical dissent. If the primary harm is communicative, then cases where the subject is unable to acquire knowledge of her experience will not count as instances of hermeneutical injustice, because the subject will have no knowledge to communicate at all. She will of course have communicative difficulties; Wood, for instance, is portrayed as *both* unable to intelligibly articulate *and* unable to acquire knowledge of her experience of sexual harassment.<sup>51</sup> But this communicative trouble comes too late. The subject has already been harmed in her capacity as a knower *before* these communicative difficulties, which are not even of the same sort required, on this interpretation, for hermeneutical injustice — with no knowledge to communicate, the communicative harm simply cannot arise in the required way. This result is problematic because we can no longer account for the most troubling central cases (namely, cases like Wood's or Sanford's) as instances of hermeneutical injustice. Interpreting the primary harm as cognitive fares no better, for as we have seen, this forces us to exclude all cases where the subject has achieved knowledge of her experience from counting as instances of hermeneutical injustice. Consequently, both cases of hermeneutical dissent and Joe's case will fail to qualify as hermeneutical injustices. We might be able to define some other species of epistemic injustice to cover

---

<sup>51</sup> Mason disputes this, arguing that the fact that Wood actively sought out a feminist consciousness-raising group with whom she could reach a better understanding of her experience indicates that she was not 'mystified by her experiences of a yet-to-be-named phenomenon' (R. Mason 2011, 297). Brownmiller's narrative is vague enough on the details to be interpreted either way. But nothing in the theoretical point about the cognitive harm turns on which reading of the case is correct. If Mason is right and Wood had a perfectly adequate understanding of sexual harassment even before that concept was coined, then we can re-imagine the case to show a cognitive harm of the kind Fricker discusses.



either interpretation's losses, but it would be more satisfying if we had a unified account given that both harms have a common cause (namely, hermeneutical marginalization).

Fricker's account is ambivalent when it comes to which of these two harms is primary. However, it is suggestive that she occasionally represents these harms disjunctively: 'the moment of hermeneutical injustice comes only when the background condition is realized in a more or less doomed attempt on the part of the subject to render an experience intelligible, *either* to herself *or* to an interlocutor' (Fricker 2007, 159, my emphasis). The disjunction suggests that the two harms are both manifestations of a more general epistemic harm. Namely, I propose that the primary harm of hermeneutical injustice is that *the subject has some distinctive and important social experience that at some crucial moment lacks intelligibility*. That crucial moment may come when the subject tries to communicate knowledge she has about her experience to an interlocutor who is ignorant or non-competent with respect to the concepts required to find her testimony intelligible. Or it may arise in the moment of the subject's experience, where her conceptual ignorance or non-competence leaves her without knowledge of that experience because it remains unintelligible even to her. So despite their apparent divergence, the communicative harm and the cognitive harm turn out to be instances of the same thing. Consequently, we can account for both hermeneutical dissent and all of Fricker's examples as instances of hermeneutical injustice.

### 2.2.3. *Global and Local Hermeneutical Economies*

Whether the primary harm of hermeneutical injustice manifests communicatively or cognitively depends on where the gap in the collective hermeneutical resource appears. That is to say, the way in which the subject's experience lacks intelligibility in the moment she suffers a hermeneutical injustice depends on which interlocutors have mastered the required concepts, and which are conceptually ignorant or non-

competent. But our account of hermeneutical injustice so far does not specify which epistemic agents and social groups have access to the required interpretive tools: a gap in the collective resource just means that at least *someone* lacks them. In order to better classify the different kinds of hermeneutical injustice we need a more precise account of hermeneutical resources that shifts our attention away from the global hermeneutical economy onto the local hermeneutical economies where dissent takes place.

In Fricker's more recent work, she begins to distinguish different types of hermeneutical injustices along these lines, using the labels 'maximal', 'midway', and 'minimal' (Fricker 2016). What is maximized or minimized is the extent of the hermeneutical gap among different social groups' and individuals' hermeneutical toolsets. In Wood's case (maximal), the hermeneutical gap is global: no individuals or groups had mastered the concept of sexual harassment. In Joe's case (minimal), the extent of the gap is extremely limited: 'he himself has no difficulty in understanding [his experience] and would easily be able to communicate to members of almost any social group', yet those to whom he most has interest in communicating — Clarissa and the police — seem not to have the required interpretive tools (166). The 'midway' band, however, is ambiguous between several different forms of hermeneutical injustice. In each case where a hermeneutical injustice occurs, a gap in the collective hermeneutical resource might represent the subject's own conceptual ignorance or non-competence, or it may represent widespread conceptual ignorance or non-competence among the members of various other social group(s) to whom she attempts to make her experience communicatively intelligible. Moreover, the idiom of maximization–minimization still glosses over details; for the spread of interpretive tools from dissenting communities out into wider circulation does not happen evenly. Understanding the kinds of hermeneutical injustice that occur in cases of

hermeneutical dissent requires further elaboration concerning these local hermeneutical resources.

Table 1 summarizes the possibilities. I have dropped permutations where the relevant concepts are in the collective resource, because that implies that everyone has mastered them. I have also dropped the permutation where the subject has not mastered the required concepts despite her own social group having them: I am assuming that if a group has the required interpretive tools, then so do all of its members. This is of course an abstraction; certainly, it takes time for concepts and mastery thereof to disperse amongst the members of a social group, and its members may disagree, developing differing conceptions. In such cases, however, it suffices to draw the line differently, circumscribing only the subgroup whose members have all mastered the relevant concepts. Note also that I am only interested in the hermeneutical resources of the subject herself and the hermeneutical resources of social groups to whose members the subject is attempting to communicate. People to whom the subject is *not* attempting to communicate may be ignorant or non-competent with respect to the relevant concepts, but this situation merely provides a necessary condition for hermeneutical injustice. Until the subject attempts to communicate with them, or they undergo the same sort of experience themselves, no hermeneutical injustice takes place. The result is six species of hermeneutical injustice.

Type of injustice	<i>Who has mastered the relevant concepts?</i>			
	Collective	Social group other than the subject's own	Subject's own social group	Subject
<i>Effacement</i>	No	No	No	No
<i>Isolation</i>	No	No	No	Yes
<i>Separation</i>	No	Yes	No	No
<i>Ghettoization</i>	No	No	Yes	Yes
<i>Exportation</i>	No	Yes	No	Yes
<i>Obstruction</i>	No	Yes	Yes	Yes

*Table 1. The species of hermeneutical injustice.*

The first scenario is what I call *hermeneutical effacement*. This form of hermeneutical injustice occurs when everyone is conceptually ignorant or non-competent with regard to the relevant concepts. Neither the subject nor anyone else will be able to acquire knowledge of her experience. Moreover, the subject and others who undergo the experience are so severely hermeneutically marginalized that they have so far been prevented from organizing any effective hermeneutical dissent. In other words, this case arises when the hermeneutical gap extends globally, beyond just the collective. We can interpret the Wood case as being of this type. Prior to the consciousness-raising seminar, there were no communities or individuals that had developed interpretive tools that fully captured the experience of sexual harassment. Moreover, the women who experienced it had not had the opportunity to come together in a supportive environment that would have allowed them to develop the needed concept. The result was that the women suffered hermeneutical injustices not just because there was a gap in the *shared* store of interpretive tools, but also because *no one* with whom they attempted to communicate had the tools required, not even themselves. Their experience of being sexually harassed thus lacked intelligibility in the moment of their experience, preventing them from acquiring knowledge about it. Therefore, Wood suffered the cognitive harm.

The second scenario is what I call *hermeneutical isolation*. This form of hermeneutical injustice occurs when a hermeneutically marginalized subject *alone* has mastered the concepts needed to make intelligible sense of her experience. That is, though the subject understands her experience well enough to achieve knowledge of it, all of her interlocutors are ignorant or non-competent with regard to the concepts needed to make sense of her attempts to share that knowledge. Moreover, her hermeneutical marginalization interferes with other agents or groups taking up the subject's interpretation of her experience, preventing her from engaging in hermeneutical dissent with any who may be similarly situated. In these situations, the

subject will not suffer the cognitive harm, because she is able to acquire knowledge of her experience, but she will suffer the communicative harm every time she attempts to communicate about that experience. We might be sceptical about whether such cases can occur as a result of a subject's lone hermeneutical dissent. As Ludwig Wittgenstein argues, a subject who were to invent (*per impossibile*) her own 'private' term or concept for a specific experience of hers would have a concept for which there was no standard of correct application, and so no standard in relation to which her applications of that term or concept would count as consistently tracking the same thing. On his view, the only way these hermeneutical tools can acquire a determinate meaning is for a community to take them up into their practices and to reach a point where the members of that community tacitly agree on public features indicating that applying those interpretive tools is appropriate (Wittgenstein 1978, §§243–275).<sup>52</sup> Arguably, it is only at this point that the subject could acquire *knowledge* of her experience using these tools. Hermeneutical dissent might begin with an individual's inchoate conceptual innovations, but in order to succeed in eliminating the cognitive harm the dissent must be brought to fruition through social processes.

However, there are cases where hermeneutical isolation can occur without hermeneutical dissent, such as when a subject becomes permanently cut off from the community in which she was brought up. For example, consider the ongoing extinction of indigenous cultures with distinctive languages and ways of thinking. Suppose the last surviving member of one such culture lives among members of the region's (now) dominant community. His practically extinct culture gave him distinctive concepts for interpreting some of his experiences — say, aspects of hunting — which his neighbours do not share, but which for him hold a distinctive practical and spiritual significance. If his neighbours do not make an effort to take up his indigenous interpretive tools, they

---

<sup>52</sup> Compare the importance of a cultural background to thick concepts, as discussed in Chapter 1, and to meaning change, as discussed in Chapter 5.

may find his interpretations of hunting to be unintelligible. In this case, the survivor's hermeneutical tools have come out of his extinct community's practices, so they will remain determinate enough for him to have knowledge of his experience. But, he will be frustrated in all his attempts to make his interpretations intelligible to his neighbours, owing to their conceptual ignorance and non-competence. He thus suffers the communicative harm. Over time, the lack of social reinforcement from members of his extinct community may also cause him to lose his grip on those interpretive tools, whittling away at his conceptual mastery until he can no longer be certain he is applying them correctly — in the end, he may suffer a cognitive harm by losing the knowledge he once had.<sup>53</sup>

The third scenario is what I call *hermeneutical separation*. This form of hermeneutical injustice occurs when the subject is ignorant or non-competent with regard to the concepts she needs make sense of her own experience, and has been unable to engage in hermeneutical dissent with members of her own social group, yet at the same time the required concepts have been mastered by members of a hermeneutically marginalized social group to which the subject does *not* belong. Once the subject comes into contact with these sympathetic strangers, her attempts to explain and understand her experience may 'click'. Those with the needed interpretive tools may recognize the subject's experience as something familiar to them, or the subject may recognize her experience in the testimony of those in the other group. The Sanford case illustrates this kind of hermeneutical injustice. As we have seen, before attending the consciousness-raising seminar, she was conceptually non-competent with regard to postpartum depression, and thus was unable to come to an accurate and

---

<sup>53</sup> A comparable fictional example can be found in Terry Pratchett's *Nation*, a low-fantasy alternative history novel about a Polynesian boy whose community is swept away by a tsunami, his interactions with Europeans and other Polynesians who come to the island, and his struggles to preserve his cultural traditions after the disaster (Pratchett 2008).

fully intelligible understanding of her own experience. At the same time, feminists, a community to which Sanford did not belong because of her political affiliations, *had* mastered this concept. Because Sanford did not have the hermeneutical resources at her disposal to acquire knowledge of her own experience, she suffered the cognitive harm until she had internalized the interpretive tools offered to her by the consciousness-raisers.

The fourth scenario is what I call *hermeneutical ghettoization*. This form of hermeneutical injustice occurs when the subject belongs to a hermeneutically marginalized group whose members have engaged in hermeneutical dissent in order to master the concepts required to attain knowledge of their distinctive experience and to communicate about it amongst themselves. But, because of this group's marginalization, no other communities have acquired such an understanding, so the subject cannot make her experience intelligible to members of other groups. One of Fricker's more recent examples captures this type of hermeneutical injustice. She draws on Mike and Trevor Phillips's account of the migration of non-white Commonwealth citizens to the United Kingdom after the Second World War (Phillips and Phillips 1998). Black colonial immigrants interpreted their experiences as a struggle to be granted their rights as British citizens, whereas white Brits often viewed the issue as one of accepting or rejecting black migrants into British society without regard for their civil rights. Fricker summarizes: 'One could say the concept of a black British citizen had not yet taken hold in white British consciousness' (Fricker 2016, 168). Because the black immigrants were frustrated in their attempts to communicate their knowledge of their civil rights to white Brits because of the latter's conceptual non-competence, the immigrants suffered the communicative harm.

The fifth scenario is what I call *hermeneutical exportation*. This form of hermeneutical injustice occurs when the subject masters concepts developed by a social group other than any of her own, and attempts to communicate with people in

her own social group who are conceptually ignorant or non-competent. The subject is thus able to acquire knowledge of her experience and to share her knowledge with members of that other social group, but not to people in her own group. Moreover, her hermeneutical marginalization hinders her attempts to import those concepts and conceptions into her local communities.<sup>54</sup> For example, we might imagine Sanford coming home from the feminist seminar and trying to explain the feminist conception of postpartum depression to her husband. From what little we know of him — namely, that he was a white, upper-class, conservative, U.S. American cis man — we would expect him to reject Sanford's new characterization of her condition. It is easy to imagine Sanford trying to explain that her depression is rooted in changes to her physiology (which are out of her control) and in social isolation (in which he is implicated), and he finding all of this unintelligible, retorting that those 'radical man-haters' have 'put ideas in her head'. When our imagined Sanford attempts to communicate the knowledge she has about her postpartum depression to her husband, she suffers the communicative harm. Such cases may generally be a prelude to a change of group membership: when people in your own community dismiss your interpretations of your own experience, the ensuing alienation is likely to spur you to leave for a more supportive group. In the end, Brownmiller tells us that Sanford divorced her husband, became a feminist activist, and 'discovered her lesbian identity' (Brownmiller 1990, 185).

---

<sup>54</sup> A worry might be raised at this point, recalling the Wittgensteinian problem for hermeneutical isolation: without being a member of the dissenting social group whose forms of life have given the required concept a determinate meaning, we might think that the subject of hermeneutical exportation cannot consistently apply that concept which she has borrowed to alleviate the cognitive harm. This worry is only half right. There will be a kind of inauthentic appropriation of hermeneutical tools, where the subject lifts a concept from an alien culture without fully understanding its meaning. In such circumstances, she will not achieve knowledge of her experience after all, because she does not really understand the other group's interpretations. But where the subject has had a more significant and ongoing interaction with the other group, it is plausible to think that she could internalize some of their concepts, and keep them on track.



The sixth scenario is what I call *hermeneutical obstruction*. This form of hermeneutical injustice occurs when the subject, her own social group, and at least some social groups to which she does not belong have all mastered the concepts required to make her experience intelligible, but there remain some social groups and individuals who remain conceptually ignorant or non-competent, and so those concepts and conceptions fail to pass into the collective resource. To count as a hermeneutical injustice, the subject must still be subject to hermeneutical marginalization: despite the relatively favourable hermeneutical climate, socially dominant interpretations still run against her and her allies. Moreover, some socially dominant groups may be actively resisting the spread of the subject's dissenting interpretations. For example, let us continue our imagined biography of Sanford. Suppose that (as may in fact be the case) after she joined the feminist community the women's health movement succeeded in exporting their understanding of postpartum depression to a significant segment of the U.S. American population. And yet, suppose the dominant cultural understanding of such experiences continues to run against their interpretations: the feminist activists, now including Sanford, are still hermeneutically marginalized *qua* women. Sanford would now be able to communicate her knowledge intelligibly to a wider range of people, but that intelligibility would continue to fail in conversation with interlocutors who have not taken mastered the concepts and conceptions of the women's health movement, especially those interlocutors who are actively resisting them. At this stage, our imagined Sanford would still be subject to hermeneutical injustice in her attempts to communicate her knowledge to some people, suffering the communicative harm.

My imagined extensions of the Sanford case also show how the kinds of hermeneutical injustices to which one is subjected might change over time. As I presented it, Sanford initially suffers hermeneutical separation; then, after internalizing the feminist conception of postpartum depression, she suffers

hermeneutical exportation; finally, after becoming a feminist herself, she suffers hermeneutical obstruction. These kinds of fluctuations occur because of changes in the subject's group membership and in different parties' hermeneutical resources. The subject may find it epistemically illuminating to expose herself to and to master the concepts in use by other social groups, possibly going so far as to change her group membership. Social groups, too, may take up concepts and conceptions from one another. In addition to the examples above, we could imagine a Sanford who never attended the consciousness-raising seminar, but whose own social group (say, upper-class Americans) eventually acquired the feminist conception of postpartum depression, bringing an end to her hermeneutical injustice without any change in her group membership. Where the subject is a member of multiple social groups, some of which are conceptually ignorant or non-competent and some of which are not, she will suffer a hermeneutical injustice when communicating with members of the former group, but not when communicating with members of the latter. So if Sanford had carried on as a conservative after becoming a feminist, and her fellow conservatives did *not* take up the feminist conception of postpartum depression, she would continue to suffer hermeneutical exportation among conservatives even though she would suffer no hermeneutical injustice among feminists.

These distinctions illustrate just how complex local hermeneutical economies can be, with diverse social groups circulating dominant and dissenting interpretive tools among individuals who may come into contact with differently situated groups. Hermeneutical dissent is not a straightforward process of inventing interpretive tools for marginal experiences and pushing them out to a wider and wider congregation. Different social groups and individuals may acquire or resist interpretive tools from other social groups, meaning that filling in collective gaps proceeds unevenly — and not, as Fricker (2016) suggests, in a straight line from the maximal form of the injustice (effacement) to the minimal form (obstruction). My account brings out these details

and allows us to fully appreciate the complexities of hermeneutical dissent while still recognizing the hermeneutical injustices that dissenting subjects suffer.

### 2.3. Wilful Conceptual Ignorance and Non-competence

The last section described some of the structural roots of widespread conceptual ignorance and non-competence, and the various kinds of hermeneutical injustices that follow. But as that discussion made apparent, the concepts that one has mastered or of which one is ignorant are not simply a matter of these structural factors. As we saw in connection with the Sanford case, in order to overcome the cognitive harms of a hermeneutical injustice, the subject may engage with people in hermeneutically dissenting social groups and master the concepts and conceptions that they have produced in order to acquire knowledge of her own or others' experiences. But we also saw that some subjects may actively resist taking up these concepts and conceptions, remaining conceptually ignorant or non-competent and thus unable to acquire the knowledge possessed by hermeneutically dissenting subjects. Such resistance to concepts and conceptions from marginal hermeneutical resources seems particularly likely to occur where the subject is from a socially dominant group and bears some prejudice against the marginalized dissenting community.

On the account I have presented, this kind of refusal to correct one's conceptual ignorance or non-competence does not, in itself, constitute a hermeneutical injustice. As Fricker observes, because hermeneutical injustice arises from a structural feature of society — the hermeneutical marginalization of social groups and individuals — no individual agents are responsible for it: 'hermeneutical injustice... involves no culprit. No agent *perpetrates* hermeneutical injustice — it is a purely structural notion' (Fricker 2007, 159). An agent who failed, because of his conceptual ignorance or non-competence, to understand the testimony of a hermeneutically dissenting speaker, and did not follow up to try to correct his ignorance or non-competence, would not be the *perpetrator* of the communicative breakdown that constitutes hermeneutical injustice

here. The root of the hermeneutical injustice here is the unfairly produced gap in the collective hermeneutical resource, and not in the hearer's failure to take up the dissenting subject's concepts and conceptions. We might think, then, that the hearer's will plays no part in hermeneutical injustice.

But this may let the conceptually ignorant or non-competent hearer off too lightly. Depending on *why* he is conceptually ignorant or non-competent and *why* he remains so after hearing testimony he finds unintelligible, we may find reason to judge that he is culpable for his conceptual ignorance or non-competence. Following this line of thought, Gaile Pohlhaus, Jr. attributes the persistence of conceptual ignorance and non-competence despite the presence of hermeneutical dissent in marginalized social groups to two factors. On the one hand, subjects not from marginalized social groups will lack the relevant social experiences, distinctive to those marginalized groups, which justify the creation of new concepts and conceptions: 'that part of the world is one to which dominantly situated knowers do not attend' (Pohlhaus 2012, 728). On the other hand, marginalized groups will find it difficult to call relatively dominant subjects' attention to the relevant experiences, because doing so often requires drawing upon the very concepts regarding which the dominant subjects are ignorant or non-competent: 'when such resources are not yet established, new epistemic resources can be preemptively prevented from calling attention to that for which they are suited, on the grounds that they do not fit anything in the experienced world' (728).

Pohlhaus claims that when relatively dominant hearers fail to take up the concepts and conceptions developed by hermeneutical dissent, 'such a refusal is *not* an inherent inability, but rather a willful act' (Pohlhaus 2012, 729). On her view, relatively dominant subjects are culpable for their conceptual ignorance, because their ignorance stems from a wilful refusal to take the experiences and epistemic resources of marginalized subjects seriously. In connection with hermeneutical injustice, Pohlhaus calls this refusal *wilful hermeneutical ignorance*, which she identifies as another kind of

epistemic injustice. Kristie Dotson argues that when such ignorance serves to maintain the conditions for hermeneutical injustice, the agent commits what she calls *contributory injustice*: ‘an epistemic agent’s willful hermeneutical ignorance in maintaining and utilizing structurally prejudiced hermeneutical resources thwarts a knower’s ability to contribute to shared epistemic resources within a given epistemic community by compromising her epistemic agency’ (Dotson 2012). For example, consider the Ford case. When Ford speaks about their genderless identity to relatively dominant hearers who have no relevantly similar experiences and are ignorant of the concept AGENDER (among others), Ford experiences a hermeneutical injustice of the communicative type. But where the hearers actively resist taking up the ideas that Ford has introduced to them, those hearers additionally commit a contributory injustice by preventing Ford from sharing those same LGBTQ+ concepts and conceptions more widely. Wilfully refusing to correct one’s conceptual ignorance and non-competence constitutes an epistemic injustice in itself, alongside the hermeneutical injustice experienced by the hermeneutically dissenting speaker.

According to Pohlhaus and Dotson, then, the conceptual ignorance and non-competence of relatively dominant subjects is the result of a wilful refusal to take up the concepts and conceptions developed by marginalized groups. However, because neither Pohlhaus nor Dotson provides an account of moral or epistemic responsibility, it is difficult to judge whether any given case of conceptual ignorance or non-competence constitutes a culpable fault. Indeed, on one popular account of moral responsibility and culpable ignorance, few agents would be culpable for their conceptual ignorance or non-competence. On this view, which I call *voluntarism*, the agent is responsible for some action only if she is both in control of her taking that action (she is not coerced, hypnotized, dissociating, etc.), and aware of the nature of her action (she knows what she is doing, who is likely to be affected, whether it is right or wrong to do so, etc.). Failing one or both of these conditions — lacking control or being

ignorant — excuses the agent from responsibility. Ignorance can fail to be an excuse when the agent is culpable for her ignorance, but this requires that her ignorance can be traced back to a clear-eyed wilfully wrongful act. The agent can be culpably ignorant only if she had taken some previous action — voluntarily and in awareness that the act was wrong — that impaired her epistemic situation at the time of the later action, such that she did not or could not know that the later action was wrong. As Michael Zimmerman summarizes, ‘all culpability can be traced to culpability that involves lack of ignorance, that is, that involves a belief on the agent’s part that he or she is doing something morally wrong’ (Zimmerman 1997, 418).<sup>55</sup>

In order to be culpable for conceptual ignorance or non-competence, on the voluntarist account, the agent must not only be in control of their refusal to take up these concepts, they must also do so despite being *aware that they are acting wrongly* — epistemically wrongly because they ignore relevant evidence (namely, the experience of marginalized subjects), and morally wrongly because they are committing an epistemic injustice. Pohlhaus’s and Dotson’s use of the term ‘wilful’ suggests that they expect this condition to hold. But I do not think an attribution of voluntary wrongdoing is plausible in a wide range of the cases Pohlhaus and Dotson are concerned with. Certainly, some subjects knowingly refuse to take up concepts from marginalized groups. Many, for example, knowingly resist taking up various concepts of sex, gender, and sexuality that have been developed in LGBTQ+ communities to make sense of the social experiences particular to people in those communities. But it is a mistake to assume that this refusal is always done in awareness that it is *wrong* to do so; indeed, such refusers typically present reasons for not taking up these concepts and conceptions. Whatever we think about the adequacy of those reasons, it is clear that relatively dominant subjects who choose to remain

---

<sup>55</sup> See also H. Smith (1983), Rosen (2003).

conceptually ignorant in this way tend to think of themselves as being *in the right*.<sup>56</sup> Moreover, we know that much prejudiced behaviour is traceable not to transparent wilful acts of wrongdoing, but rather to evaluative judgements, biases, and prejudices that operate at a more obscure, implicit level.<sup>57</sup> Seen in this light, the range of cases where conceptual ignorance or non-competence are *wilful* seems small indeed.<sup>58</sup>

Perhaps this is too quick. It may be that the conceptually ignorant or non-competent hearer's failure to take up marginal concepts and conceptions stems from something else for which the hearer is culpable, in such a way that she is also culpable for remaining ignorant or non-competent. We need only trace her refusal to take up the relevant concepts and conceptions to one or more wilful acts of wrongdoing in her past that impaired her present epistemic situation such that she does not know *now* that it would be wrong to refuse to take up those concepts and conceptions. Perhaps we can trace this refusal to some earlier wilful decision to wrongfully ignore the testimony of marginalized subjects with regard to their experiences or to take too seriously views opposed to marginal accounts, such that the hearer now does not know now that her refusal to take up the relevant concepts is wrongful. Unfortunately, for the same reasons given above, I think it is unlikely that we will find many instances where this condition is fulfilled. Again, I doubt that many agents who commit acts that later impair

---

<sup>56</sup> Many agents who are conceptually ignorant or non-competent with regard to concepts produced by oppressed groups are likely to be from socially conservative groups who are themselves disadvantaged. Here, it is important to bear in mind that the dominance of a social group is always relative to other social groups, and potentially highly localized to particular interactions.

<sup>57</sup> See Brownstein and Saul (2016a, 2016b) for articles on implicit bias. See also Fricker (2007, 60–85), for an argument that prejudice distorts the subject's ability to perceive moral and epistemic reasons at an unreflective level. On the epistemology of prejudice, see Begby (2013).

<sup>58</sup> Indeed as Zimmerman (1997) foresees, voluntarism leads to the conclusion that almost no one is culpable for their racism, sexism, or other prejudices. He is willing to bite the bullet on this result, but I think it reveals voluntarism to be seriously flawed because it is almost absurdly strong — indeed, Rosen (2004) shows that voluntarism may be used to motivate skepticism about moral responsibility. In Chapter 3, I consider alternatives to voluntarism.

their epistemic situation in this way do so in awareness that they are acting wrongly; in many cases, they seem more likely to think that they are acting rightly. And again, these previous actions could be explained by the same kinds of implicit judgements, biases, and prejudices that explain many instances of failing to take up concepts from marginalized groups in the absence of any awareness that this failure is wrongful.

Another option to try expanding the range of cases where the subject is culpable for his conceptual ignorance or non-competence would be to appeal to the subject's vicious character traits. As Elinor Mason and Alan Wilson argue, a great deal of cultural ignorance — that is to say, widespread ignorance of moral facts in a given cultural community — may be attributable to widespread moral-epistemic vices of arrogance and laziness (E. Mason and Wilson 2017). We might think that widespread conceptual ignorance and non-competence stemming from the hermeneutical marginalization of various social groups is sustained through the same vices. Medina (2013), for instance, claims that subjects from relatively dominant social groups are more likely than differently situated subjects to possess these kinds of vices — arrogance and laziness, as well as closed-mindedness. These vices might manifest themselves in a failure to take on board hermeneutically dissenting concepts and conceptions when a marginalized speaker uses them.

This strategy seems more promising, as unwitting wrongdoing attributable to one's vices intuitively seems to be culpable despite not stemming from a wilful decision to impair one's ability to know that one's action is wrong. However, we do not yet have an account of why ignorance resulting from moral or epistemic vices is culpable. And on the voluntarist account, this will just prove to be another non-starter. Accordingly, in the next chapter, I turn to a defence of an alternative view of moral responsibility. We will then be in a position to judge whether the agent is responsible for her conceptual ignorance or non-competence — as well as for the use of wrongful concepts and conceptions — in a range of scenarios.



#### 2.4. Summary

In this chapter, I presented the notions of conceptual ignorance and conceptual non-competence as the lack of a particular concept in one's repertoire, and the lack of the appropriate knowledge required to have mastered a particular concept or the possession of a mistaken conception. I explored some of the social-structural sources of conceptual ignorance and non-competence in connection with hermeneutical marginalization. This discussion also revealed that widespread conceptual ignorance and non-competence may be connected to various forms of hermeneutical injustice. These injustices may harm marginalized subjects in their capacity as knowers by preventing them from acquiring knowledge of their own distinctive and important social experiences, or by preventing them from sharing the knowledge they do have of those experiences. I then turned back to conceptually ignorant or non-competent subjects to see whether we can find fault with their behaviour beyond their connection to hermeneutical injustices. While it appears that these subjects may be culpable at least some of the time for their conceptual ignorance or non-competence, we do not yet have a satisfying explanation for why.

We might collect the use of wrongful concepts and conceptions (discussed in Chapter 1) and conceptual ignorance and non-competence (discussed in this chapter) under the label *conceptual incapacities*. In the next chapter, I defend an alternative account of moral responsibility that provides conditions under which the subject may be culpable for her conceptual incapacities.



### 3. Responsibility for Conceptual Incapacities

In the previous two chapters, I outlined three kinds of what we may call *conceptual incapacities*.<sup>59</sup> In Chapter 1, I discussed thick ethical and social concepts the use of which may be problematic morally, epistemically, or socially. In Chapter 2, I introduced the notions of conceptual ignorance (not possessing a concept) and conceptual non-competence (possessing a mistaken conception or lacking the range of knowledge or other cognitive requirements for mastery of a concept), both of which are implicated in hermeneutical injustice and associated harms. At the end of Chapter 1 (when discussing an objection from moralism), and again at the end of Chapter 2, I noted the possibility that individual agents may be *culpable* in instances where their conceptual incapacities produce the moral, epistemic, or social problems I described. This suggests that we may be *responsible* for our conceptual incapacities. But to see whether we are, we need an account of responsibility on the table.

In this chapter, I argue that we are sometimes responsible for our conceptual incapacities. The first six sections are mainly concerned with outlining a theory of responsibility. §3.1 introduces a distinction between responsibility as attributability, as accountability, and as answerability. I argue that attributability is the fundamental sense with regard to the notion of being responsible for something. §§3.2–5 then examine several prominent accounts of responsibility attribution, which I call voluntarism, the deep self view, the rational relations view, and psychologism. I comment briefly on how each of these views handles the issue of conceptual responsibility, but express doubts about the ability of each account to deliver intuitive results in various standard cases of responsibility. §3.6 sketches my preferred view of

---

<sup>59</sup> This term follows a suggestion from Matthew Congdon: ‘alongside... “conceptual capacities” we could introduce the notion of “conceptual incapacities.” This would refer not simply to a potential yet lacking conceptual capacity in one’s conceptual repertoire, but to a substantive locus of false belief and ignorance in one’s conceptual repertoire’ (Congdon 2015, 90).

responsibility, which I call reasonable psychologism. §3.7 then applies this view to the issue of conceptual responsibility in more detail.

### 3.1. Three Senses of Responsibility

'Responsibility' and related terms are notoriously polysemous in both everyday language and in moral philosophy. First, let me discard those senses that I will not be using. One sense 'responsibility' and 'responsible' can take is a purely causal notion, as when we say things like, 'The tsunami was responsible for millions of dollars in damage and hundreds of deaths'. But when it comes to our actions, merely giving a causal explanation for why some state of affairs occurred as a result of an agent's movements in the world does not yet establish anything morally significant about the agent's involvement in that state of affairs. There are also multiple senses of 'responsibility' in legal or professional contexts. But while importantly connected to moral responsibility in various ways, jurisprudence and professional conduct are not my concern here.<sup>60</sup> Another sense that 'responsible' may take is a label for a specific ability or capacity, as in the contrast between a 'responsible agent' and a 'non-responsible animal'. While an important precondition for moral responsibility, the nature of this agential capacity is not what I am directly concerned with. Yet another sense 'responsible' may take is a moral, epistemic, or professional virtue, as in 'a responsible person', 'a responsible inquirer', or 'a responsible administrator'. While this use has moral significance, I suspect it is ambiguous between sets of discrete virtues and capacities, and describes a general trait of the agent rather than the agent's moral status with respect to a particular action of hers. Finally, sometimes 'responsibility' is taken to mean a duty or obligation or commitment, assumed or imposed, as in 'It's my responsibility to cook, so it's your responsibility to wash up', or 'I'll be responsible for the children's lunches', or

---

<sup>60</sup> Indeed, the possibility of legal enforcement of a conceptual scheme was the worry raised by the 'thought police' objection in Chapter 1.

‘We have a moral responsibility to ensure that every citizen has access to health care’. While plainly in the realm of the moral — or at least the interpersonal — this sense of the term is pointing to a different aspect of ethics that is concerned with what agents are required to *do*, independent of which of their actions they are responsible *for*. While I discuss assumptions of responsibility and obligations we may have in response to being responsible for our conceptual incapacities in Chapter 4, I have avoided using ‘responsibility’ as a synonym for ‘obligation’ for the sake of clarity.

Having set these aside, there are three significant senses of ‘responsibility’ and ‘responsible’ that have been identified in the recent literature. I will use the taxonomy employed by Andrew Eshleman (2016) and David Shoemaker (2011), which distinguishes between responsibility as *attributability*, as *accountability*, and as *answerability*.<sup>61</sup> The terminology varies in the literature, with each of the three terms slipping between one another in different philosophers’ usage. As will become clear, however, each refers to a distinct aspect of moral responsibility, though the conditions for each are jointly satisfied in the central cases. Below, for the sake of consistency, I have edited quotations as appropriate to match my usage.

In the *attributability* sense, the agent is responsible for X just in case X is properly attributable to her, such that the agent is open to moral appraisal with regard to X. Sometimes this is put in terms of a metaphorical moral ledger, where good actions that are attributable to the agent assign a credit and bad actions attributable to the agent assign a debit.<sup>62</sup> What makes a credit or debit legitimate is the presence of an appropriate relation between the agent and her behaviour. This relation is often put in terms of a connection between the agent’s taking the action and the agent’s judgements, desires, commitments, responsiveness to reasons, or some other aspect of

---

<sup>61</sup> This taxonomy draws upon Gary Watson’s earlier distinction between attributability and accountability. See Watson (1996).

<sup>62</sup> See Feinberg (1970), Glover (1970), Zimmerman (1988)

her psychology. One analysis holds that the agent's actions must have been voluntary, that is to say, the agent must have been in control of her taking that action and aware of the moral character of that action (i.e. whether it is morally right or wrong).<sup>63</sup> Others require only that the agent's behaviour must express her evaluative judgements, level of moral concern for those affected, or responsiveness to the ethically relevant information that is available to her.<sup>64</sup> Still others argue that the agent's behaviour must stem from values that she has chosen to identify herself with — the evaluative commitments that led her to take that action must be endorsed at the level of her 'deep self'.<sup>65</sup> The weakest form of the attributability condition is simply that the agent's behaviour was the cause of something — it is enough that she did it for her to be responsible for it.<sup>66</sup>

In the *accountability* sense, the agent is responsible for X just in case she is an appropriate target of some kind of third-personal moral response for X. Those who defend this interpretation of moral responsibility are principally influenced by P. F. Strawson's view that an analysis of moral responsibility should begin by examining the practices through which we socially reward or punish agents for their morally salient behaviour (Strawson 2008).<sup>67</sup> These moral responses include moral reactive attitudes, such as praise or blame, as well as practices of punishment or reward, and related moral emotions such as resentment, disappointment, admiration, and so on. The thought is that we can sidestep difficult problems concerning traits about the agent, such as her psychological states or whether she is metaphysically free, or facts about the agent's situation, such as the effects of random chance on the outcome of her action

---

<sup>63</sup> See, again, H. Smith (1983), Zimmerman (1997), Levy (2014).

<sup>64</sup> See A. M. Smith (2005), Arpaly (2002, 2006), Adams (1985).

<sup>65</sup> The originator of this kind of view is Frankfurt (1971). Other accounts that have been identified as deep self views include those presented by Watson (1975) and Taylor (1976).

<sup>66</sup> Williams (1981) and Adams (1985) might be taken to hold such views.

<sup>67</sup> See also Darwall (2006), Shoemaker (2007, 2017).

or the cultural circumstances in which she finds herself, by turning to our actual practices of holding people accountable. Moreover, social practices of expressing these attitudes are fundamental to how moral communities are organized.

In the *answerability* sense, the agent is responsible for X just in case she is open, in principle, to a moral demand that she take some further action that acknowledges her connection to X. The kind of demand that is nearly always emphasized in the literature is a demand that the agent ‘explain herself’, that is to say, *provide her reasons* for taking the action she is responsible for. Another way of putting the point is that the agent is open to questions of the form, ‘Why did you do X?’, where this question is a request for the agent’s ethical justification for X.<sup>68</sup> Taking a broader view of this sense of responsibility, it has also been argued that the more central kind of demand is for the agent to *apologize* for what she has done.<sup>69</sup> To this I would add that agents who are responsible for some action are routinely expected to take various other actions to *answer for their actions*.<sup>70</sup> Sometimes, the agent is morally required to compensate the victims of her actions or to otherwise make amends for the damage caused. In other cases, the agent is morally required to undertake a project of moral self-improvement to prevent similar wrongdoing in the future. It may be that the agent is morally required to pay some penance or take some action to atone. Exactly what is required will depend upon the nature of what the agent has done and the severity of the moral wrong. For minor acts of wrongdoing it may suffice to apologize or to take some action that symbolizes an apology, such as offering a gift; for severe moral infractions, it is possible that the agent will spend the rest of her life trying to make up for what she has

---

<sup>68</sup> See Oshana (1997), Scanlon (1998), A. M. Smith (2000, 2005, 2012), Hieronymi (2014).

<sup>69</sup> See Enoch (2012). The term he uses is ‘accountability’, but he explicitly distances his account from the Strawsonian reactive-attitudes sense of ‘responsibility’, and moreover takes himself to be expanding upon Oshana’s view. See also fn. 71 below.

<sup>70</sup> Here I am departing from the standard use of the term ‘answerable’. I am not certain what other term would be appropriate here.

done. So, as I will use the term, in this broader sense *answerability* refers to the sense in which the agent is responsible for X just in case she is morally required to respond to X in some way that acknowledges her connection to X. These responses include offering explanations or justifications for X, as well as apologizing for, feeling bad about, and trying to make amends for X.

Various philosophers have argued that each of these senses — attributability, accountability, and answerability — is the *fundamental* analysis of the concept of moral responsibility. Strawson and his followers, for example, argue that to be accountable for an action — to be the proper target of a moral reactive attitude such as blame — just is to be responsible for that action. By contrast, Marina Oshana contends that being answerable for one's action — being open to demands to provide one's reasons for taking that action — is the fundamental condition for ascribing responsibility for that action (Oshana 1997).<sup>71</sup> But these kinds of approach conflate the notion of *being* responsible for an action with *holding* someone responsible for her actions and *taking* responsibility for one's actions. Of the three senses of 'responsibility', only attributability views provide the conditions for being responsible for something; an account that is based around accountability or answerability will always need to be supplemented with an attributability condition. Any satisfactory account of *why* the agent is an appropriate target of a reactive attitude for X or open to demands to provide her reasons for X (or to take some other action in recognition of X) must explain what it is about the agent's connection to her action that makes her open to these responses. And to explain this connection between the agent and her action just is to provide an attributability condition — something that qualifies the action as being

---

<sup>71</sup> Oshana's term is 'accountable', but she uses it in contradistinction to Strawsonian reactive-attitudes views. She occasionally uses the term 'answerable' in her paper, but she prefers 'accountable' in order to avoid confusion with Kurt Baier's sense of 'answerable', which is different again from what I have outlined here. See Oshana (1997, 82 n. 17) and Baier (1987).



to the credit or discredit of the agent. So the notion of *being* responsible for something is centrally concerned with attributability.

The notions of accountability and answerability, on the other hand, are better associated with the practices of holding someone responsible for something and taking responsibility for something, respectively. With respect to *holding* someone responsible for what they have done, this notion typically refers to blaming or punishing the agent for her actions — a subset of the moral responses captured by accountability. And with respect to *taking* responsibility for one's actions, this notion usually refers to 'owning up' to one's actions (that is to say, acknowledging fault and admitting that one's reasons for acting in such a way were not enough to justify it) and then taking some further action to make up for what one has done — and these are just the actions that one is morally required to take in virtue of being answerable, in my sense, for one's action.

Moreover, accountability and answerability can also occur independently of there being an action attributable to the agent. These situations occur notably in connection with the range of cases David Enoch calls 'penumbral agency': where the agent's actions have morally bad outcomes due to bad luck, or people with whom the agent has a close relationship do something morally wrong, or collectives of which the agent is a member do something morally wrong (Enoch 2012). Pamela Hieronymi refers to a similar but more limited range of cases under the label 'jurisdictional responsibility': 'things you can affect and so perhaps control through your intentional actions, together with the fact that... you are rightly expected to affect and control them in certain ways' (Hieronymi 2014, 10). In these kinds of cases, it is not unusual to be accountable or answerable for events that are not attributable to one's agency at all

(‘I’m holding you responsible for your son’s behaviour’; ‘You need to take responsibility for your country’s actions’).<sup>72</sup>

One potential reason why others have attempted to analyse the notion of being responsible as fundamentally to do with accountability or answerability may be that in the central cases of being morally responsible for something, the three senses of ‘responsibility’ coincide. To be precise, when the agent is *at fault* for some action (or, equivalently, when she is *culpable* for some action), that action is attributable to her, she is blameworthy for that action, and she is morally required to answer for that action. Consider this example. Wade and Peter are friends having a heated argument. In the course of the exchange, Wade punches Peter in anger. Suppose that Wade has no excuse for this violent overreaction. His act of punching Peter was taken voluntarily and in awareness that it would be wrong to do so, so on at least one account, it is properly attributable to him. Furthermore, Wade is also a proper target of blame, both from Peter and potentially from anyone else; he has acted wrongly in punching his friend and deserves to be reproached. Further still, Wade is open to questions like ‘Why did you hit Peter?’ (to which he will not be able to offer a justificatory reason; that he was angry is not a good enough reason for assaulting his friend). Finally, Wade is under a moral obligation to apologize to Peter for lashing out physically; he may additionally owe it to Peter to take some other actions to show that he is sorry for what he has done. Because Wade is at fault for punching Peter, he is accountable and answerable for this action, which is properly attributable to him.

My purpose in this chapter is to give an account of the conditions under which we are responsible for our conceptual incapacities. Since this aim is concerned with *being* responsible for something, I concentrate on the conditions for attributability. In §§3.2–5, I go over several candidate theories of attribution conditions, which I call

---

<sup>72</sup> I will come back to penumbral agency generally and in connection with conceptual incapacities in the Chapter 4.

voluntarism, the deep self view, the rational relations view, and psychologism. In each case, I comment on whether the view allows that we may be responsible for our conceptual incapacities; but, in the course of the discussion, problems with each of these views emerge. In §3.6, I outline my preferred account, which I call reasonable psychologism, and in §3.7 I discuss responsibility for conceptual incapacities on this preferred account in more detail. In Chapter 4, I discuss the possibility of being answerable for our conceptual incapacities even when our use or ignorance of certain concepts is not properly attributable to us. In Chapter 5, I expand on the nature of the obligations we have in virtue of being answerable for our conceptual incapacities, and briefly touch on mechanisms by which we may be held accountable for them.

### 3.2. Voluntarism

According to *voluntarism*, to be responsible for something, the agent must have voluntarily chosen to bring it about. The agent is morally responsible for an action if and only if she was freely in control of that action, and was aware of the moral worth of the action (i.e., whether taking that action would be morally right, wrong, or neutral). If either the control condition or the epistemic condition fails to be fulfilled — for example, if the agent is coerced or mind-controlled or psychotic when she acts, or if she is ignorant that it would be wrong to take that action — then the agent is excused from responsibility for that action. But excuses are also defeasible, should the agent be culpable in turn for lacking control or awareness. With regard to ignorance, Gideon Rosen sums up this position as follows: ‘Whenever an agent acts from ignorance, whether factual or moral, he is culpable for the act only if he is culpable for the ignorance from which he acts’ (Rosen 2003, 64). Voluntarism can then be applied recursively to cases of ignorance, as Zimmerman explains: ‘all culpability can be traced to culpability that involves lack of ignorance, that is, that involves a belief on the agent’s part that he or she is doing something morally wrong’ (Zimmerman 1997, 418). Holly Smith develops this view of culpable ignorance around the notion of a ‘benighting

act': some deliberate action or omission on the agent's part which impaired or failed to improve the agent's cognitive position, such that the agent is unaware or cannot know that the action she is *now* taking is wrong. Moreover, 'the benighting act must be objectively wrong' (H. M. Smith 1983, 547). The agent *could* have been (or made it possible to be) aware of the wrongfulness of her action, and *ought* to have done so because of the future moral risk, but did not. And, to transfer culpability via ignorance to the wrongful action done in ignorance, the benighting act 'must also be one for which the agent is *culpable*' (548), meaning that the agent must have also been aware that impairing her cognitive position was wrong. As Rosen notes, this recursion may repeat as many times as necessary until we either arrive at an excuse (some non-culpable instance of ignorance) or a wilful act of wrongdoing (Rosen 2004).

As noted in Chapter 2, voluntarism makes it very difficult to conclude that anyone is culpable for their conceptual ignorance or non-competence, because it is highly doubtful that these states may be traced to some wittingly wrongful act. It is just as doubtful that the use of problematic concepts will meet the requirements of voluntarism. When we acquire propositional attitudes, our concepts are typically deployed through cognitive processes over which we have no direct control. And in reflective moments when one consciously controls one's thinking and chooses to use one concept rather than another, it is rare to deliberately select a concept that one believes to be problematic in some way. As was the case with agents who dismiss concepts produced through hermeneutical dissent, those who use problematic concepts in their reflective thinking usually take themselves to be in the right, either unaware that the concept is potentially wrong to use, or in disagreement with those who claim that it is wrong to use. Moreover, again, it is doubtful that their ignorance of wrongdoing traces to some clear-eyed act of wrongdoing. If voluntarism permits any instances of responsibility for our conceptual incapacities, it is only for a vanishingly small range of cases.

However, these difficulties are not specific to the question of responsibility for our conceptual incapacities. Voluntarism is a very strong view — indeed, implausibly so. For, it entails that very few intuitive cases of acting from culpable ignorance meet the conditions for responsibility. Consider the following example:

*Peanuts.* Chuck is organizing a party at his home. One of his invited guests is his childhood friend Lucille, who is seriously allergic to peanuts. Lucille has told Chuck about her allergy multiple times in the past, but it has slipped his mind — it has been a while since she has explicitly mentioned it, but she quite reasonably assumes that good ol' Chuck will remember. He does not. Moreover, Chuck does not think to ask Lucille if she has any dietary restrictions; it simply does not occur to him to ask, since, like most of us, he believes that he would remember anything this important about his closest friends. Among the snacks he leaves out for his guests are mixed nuts that contain peanuts. At the party, Lucille comes into contact with some peanuts, and has an anaphylactic reaction, ending the night in the hospital.

Intuitively, there is a very strong pull towards the judgement that Chuck is culpable for putting Lucille's life in danger. It seems obvious that he *should have remembered* what she told him about her allergy; his ignorance that putting out peanuts would be wrong does not seem to excuse. But voluntarism does not enable us to conclude that Chuck is culpable for his ignorance: he has not knowingly done anything to make himself forget about Lucille's allergy. Indeed, it is difficult to imagine what he *could* knowingly have done to make himself ignorant, barring science fiction scenarios (a memory-wipe) or significant substance abuse (literally drinking to forget). And because Chuck is not culpable for forgetting about Lucille's allergy, he is not culpable for putting her life in danger, either. Voluntarism provides no way to conclude that, in the vast majority of cases, we are culpable for forgetting morally important facts — and therefore, the view provides no way to conclude that we are culpable for acting wrongly out of such ignorance.

George Sher provides an array of further intuitive counterexamples to voluntarism (Sher 2009). He calls voluntarism 'the searchlight view', picturing the agent's conscious awareness as a narrow beam illuminating some but not all of her

actions. The first class of counterexamples is lapses of attention, such as a ferryman who becomes distracted by reverie until it is too late to pilot his ship around a reef. The second is failures of judgement, for example, a woman who hears the sound of someone in her home, grabs a loaded gun, and fires it upon seeing an unexpected person she judges to be a dangerous intruder — who turns out to be her son who has come home early for the holidays. The third is certain instances of false belief, for example, an American anarchist who as a result of his political convictions thinks that capitalism is so evil that the lives of individuals can excusably be sacrificed in order to ensure the coming revolution, and thus believes (falsely) that his killing a security guard in the process of a bank robbery (to procure funds for the revolutionaries) is morally permissible.

In each of these cases, Sher attributes the agent's ignorance to some combination of her attitudes, dispositions, and traits (more on this in §3.5). But it is not plausible that, in these cases, the agent has committed a benighting act, knowingly making herself ignorant while fully aware that to do so would be morally wrong. As Sher observes, 'we rarely make decisions with the clear understanding that they will cause us to acquire traits or habits that are markedly worse than normal' (Sher 2009, 38). Lapses of attention are often involuntary; they usually just happen, without any previous ill will. A badly-made judgement is not usually made in awareness that the judgement is wrong, and in most cases will not result from an earlier wittingly wrongful decision to impair one's cognitive position. Ignorance need not come from a conscious decision to shirk one's moral-epistemic obligations. Yet, Sher claims, in each of these cases the agent seems, intuitively, to be culpable both for her ignorance and for her actions taken out of ignorance.

Zimmerman anticipates these worries about ignorance stemming from the agent's moral or psychological traits or dispositions. In a brief discussion of sexism and racism, he notes that, on his view, 'responsibility for these vices and the vicious

behavior in which they issue is incurred less frequently, perhaps far less frequently, than is commonly supposed' (Zimmerman 1997, 425). He is willing to bite the bullet, for, he claims, not being culpable for their beliefs and actions does not save vicious agents from disapprobation. On his view, the range of moral evaluation is broader than praise, blame, and other responses that depend on the agent's being responsible for her actions. So the voluntarist, Zimmerman suggests, can still criticize agents for their moral character defects, for their morally wrong actions taken in non-culpable ignorance, and for their problematic conceptual capacities, so long as his disapproval does not involve any attribution of responsibility for these traits. In effect, on the taxonomy I introduced in the previous section, Zimmerman argues that we should view agents who act wrongly out of ignorance as accountable (open to some form of reproach distinct from blame) and answerable (open to critical demands that they revise their views), but not *at fault* for their actions. However, it is unclear just what kind of criticism or opprobrium Zimmerman has in mind.

Furthermore, while I am sympathetic to the idea that attributability, accountability, and answerability may come apart, I think Zimmerman's suggestion gives up too much. The range of cases where we are, according to the voluntarist, non-culpably ignorant, *extends to the vast majority of cases of ignorance*. It is rare that we knowingly impair our cognitive situation in awareness that to do so is morally wrong because it will make it difficult to see that some future range of actions is wrong. Indeed, Rosen argues from a kind of voluntarism to a sceptical position on the vast majority of cases where we ordinarily judge that the agent is culpable. Rosen starts from the principle that to be culpable for a wrongful action taken in ignorance the agent must be culpable for her ignorance. To be culpable for her ignorance, the agent must be culpable for some prior culpable failure to fulfil a moral-epistemic obligation to inquire, to reflect, or to do whatever else that would have prevented her ignorance. If this moral-epistemic negligence was itself because of the agent's ignorance, then *that*

ignorance must be the result of some culpable failure to fulfil a moral-epistemic obligation. From this, Rosen argues that ‘every culpable bad action must be the causal upshot of a genuinely akratic act or omission’ (Rosen 2004, 307, italics removed). That is to say, the only cases where the agent is straightforwardly (not derivatively) culpable for some act of wrongdoing are those where the agent knew that what she was doing was wrong. But, Rosen argues, ‘given the opacity of the mind — of other minds and even of one’s own mind — it is almost always unreasonable to place significant confidence in... a judgment [that the agent was akratic]’ (308). And so, we have reason to withhold judgement with regard to *every* putative case of culpability: for, to be culpable, the agent must either have been akratic or have had some ignorance traceable to an akratic act or omission. Starting from voluntarism, we arrive at the conclusion that few, if any, of our judgements of culpability are justified. We end up in a position where responsibility cannot reasonably be attributed in nearly any case of wrongdoing, and not just in a more limited range of cases as Zimmerman seems to suggest.

I take it that this endpoint is intolerable. If Rosen is right, then voluntarism is self-defeating as a theory of moral responsibility, because although it maintains space for being *objectively* culpable for one’s actions, it is impossible for us to *know* whether anyone ever meets these conditions. The upshot is that even though voluntarism is a coherent account of attributability, it cannot be used to ground a judgement regarding the agent’s openness to blame or moral demands, making responsibility attribution an inert and superfluous notion. But even if we do not accept Rosen’s sceptical argument — say we reject his claim that the mind is ‘opaque’ (cf. FitzPatrick 2008) — voluntarism is sufficiently problematic to warrant rejecting it. As I argued above, voluntarism leads to the conclusion that *many* intuitive cases of culpability are in fact excused: Chuck is excused for putting Lucille’s life in danger because he forgot, the agents described by Sher are off the hook because they variously did not voluntarily



lack awareness of their moral errors, and people with vices such as sexism or racism are not responsible because they did not voluntarily do anything to acquire those traits. This is a significant range of cases, suggesting that we would do well to re-examine the attributability conditions that led to this unsatisfying result.

### 3.3. The Deep Self View

One alternative to voluntarism is known as the *deep self view*, an influential version of which originates in the work of Harry Frankfurt. He makes a distinction between first-order and second-order desires, which correspond, respectively, to the superficial self and the deep self. First-order desires, understood broadly, are the agent's wants, goals, motivations, values, and so forth. Those first-order desires that are effective in getting the agent to act — even if the agent has other, conflicting first-order desires at the time — constitute the agent's will, or superficial self. As Frankfurt writes,

Thus the notion of the will is not coextensive with the notion of what the agent intends to do. For even though someone may have a settled intention to do *X*, he may nevertheless do something else instead of doing *X* because, despite his intention, his desire to do *X* proves to be weaker or less effective than some conflicting desire. (Frankfurt 1971, 8)

However, being reflective creatures, human beings are not subject to our first-order desires alone. Second-order desires reflect the agent's reflective perspective upon her first-order desires: what she desires to desire, the goals she wants to have, the motivations she would like to have, the values she wishes to hold. Those second-order desires that concern the contents of the agent's will constitute her deep self, reflecting 'who she really is'. On Frankfurt's view, for the agent to be responsible for her actions, she must not only have freedom to act, that is to say, the ability to act as she wills, or equivalently, to follow her first-order desires. In addition, to be responsible for an action she takes, the agent must also have accepted the first-order desires that

motivate that action as her own.<sup>73</sup> That is to say, the agent's first-order and second-order desires must be in alignment: she must want to have the desire that motivated the action that she takes. To be responsible for something, it must reflect the agent's deep self, and not merely her superficial self. Frankfurt illustrates with the example of a drug addict. Because of a physiological dependence on the drug, the addict has a strong first-order desire to use the drug, but reflectively desires not to have the desire to use the drug. He struggles to overcome his first-order desire to use the drug and tries various means to overcome it, to no avail: 'He is an unwilling addict, helplessly violated by his own desires' (12). This agent's addiction is not a part of who he 'really' is, because he does not accept the relevant first-order desires; in this area of his life, his will is not his own. Therefore, he is not responsible for his addiction.

The deep self view could potentially expand the range of cases where the agent is culpable for her conceptual incapacities. Where the agent has reflectively endorsed or accepted her use of certain concepts, she may be responsible for her use of those concepts. Where those concepts are problematic in the ways I have discussed, the agent would then be culpable for using them. Where the agent has reflectively endorsed or accepted her use of a mistaken conception, she may be culpable for her conceptual non-competence. The idea of reflectively endorsing one's ignorance on some matter, whether of a concept or anything else, may strike us as strange, but we can easily imagine an agent whose prejudice (conscious or unconscious) against a marginalized group leads her to a desire not to engage with that group's interpretations. Should she reflectively accept this desire, she may be responsible for her ignorance of the concepts developed by that group's hermeneutical dissent.

---

<sup>73</sup> Frankfurt sometimes writes of *endorsing* the contents of one's will, but his later work shows a preference for the language of *accepting* the contents of one's will. His reason for the shift is that the notion of endorsement implies a positive valuation, while acceptance is neutral in that respect. That is to say, the agent may want to have certain desires without necessarily thinking that these are good desires. See Frankfurt (2002, 20 n. 2).

But these ascriptions of culpability for conceptual incapacities presuppose that the deep self view can give us an account whereby the agent may be responsible for acting out of ignorance. For in each of the cases just mentioned, it is likely that the agent's reflective endorsement of her conceptual incapacities will be done in ignorance of the moral, epistemic, and social problems associated therewith. To be culpable for conceptual incapacities endorsed at the level of her deep self, the agent would have to be culpable for her ignorance that these incapacities are bad things to have. And on the face of it, the deep self view does not provide a satisfactory account of culpable ignorance. Recall that, on this view, to be responsible for an action, the agent must accept the first-order desires that motivate that action; that is to say, she must have a second-order desire to be motivated to take that action. To be culpable for her ignorance and thus for her unwitting wrongdoing, the agent's ignorance must have been caused by one of her actions that was motivated by a desire that she accepts.

But notice that we have just been led back to voluntarism. For on the deep self view, culpable ignorance must be caused by a benighting act as well; all that the deep self view adds is that the first-order desire motivating the benighting act must be a desire that the agent wants to have. All of the same problems that we encountered with regard to voluntarism will thus recur. In *Peanuts*, for example, Chuck will still be found to be non-culpably ignorant. He took no action to induce his forgetting about Lucille's allergy and had no first- or second-order desires regarding his ignorance. To repeat, we usually forget things without consciously taking an action to bring it about, much less from taking an action motivated by desires we have reflectively accepted. The agent's acceptance of her conceptual incapacities may be due to ignorance that she has not reflectively accepted, but which intuitively still seems culpable, for instance, ignorance rooted in a prejudice of which the agent is unaware.

Moreover, there will still be a wide range of cases where the agent has not reflectively accepted her conceptual incapacities, yet still seems to be culpable. It

seems that an agent who has never reflected about her use of a derogatory concept like SAVAGE, for example, should still be culpable for this conceptual incapacity. Requiring that her deep self ‘sign off’ on this problematic concept seems unnecessary. More disconcertingly, the deep self view leads us to the conclusion that the agent is off the hook for bad behaviour of hers that she has reflectively *rejected*. But there seem to be cases where the agent feels alienated from and wishes to reject certain desires of hers while still being fully culpable for what she does. Consider cases of akrasia. In these scenarios, the agent knowingly acts wrongfully because she succumbs to a strong first-order desire to do so, despite having a second-order desire not to have that first-order desire. On the deep self view, this would excuse the akratic agent from responsibility, but we do not ordinarily judge that akratic agents are non-culpable. Indeed, they seem to be among the clearest cases of fault. With regard to our conceptual incapacities, we can easily picture a case where the agent routinely finds herself using concepts that she has rejected at the level of her deep self — despite having realized that thinking of indigenous people as ‘savages’ is wrongful, for example, she may continue to do so where she does not deliberately remind herself not to. Given that such an agent may be engaged in a difficult process of cognitive improvement, we may wish to hold back from blaming her to the same degree as the agent who has reflectively accepted this problematic way of thinking. But to say that she is not *responsible* simply because she disapproves of what she has done seems mistaken.

### 3.4. The Rational Relations View

An alternative view of responsibility attribution that rejects the assumptions of voluntarism has been defended by Angela Smith.<sup>74</sup> On her account,

---

<sup>74</sup> Similar views are held by Oshana (1997), Scanlon (1998), and Hieronymi (2014). While this family of views considers answerability to be the fundamental form of responsibility, Smith’s rational relations account aims to provide an attributability condition that grounds answerability.

An agent is responsible for  $\Phi$  just in case  $\Phi$  bears a rational connection to the agent's evaluative judgements... to say that an agent is morally responsible for some thing is to say that the agent is open, in principle, to demands for justification regarding that thing. (A. M. Smith 2012, 577-78)

Because of the importance of these rational connections between what the agent is responsible for and the agent's evaluative judgements, Smith calls this the *rational relations view*. By the agent's 'evaluative judgements', Smith has in mind something much broader than the agent's reflectively endorsed judgments of what is valuable or of which considerations count in favour of acting in one way or another. The relevant notion of evaluative judgements is expansive, encompassing any of our 'tendencies to regard certain things as having evaluative significance' which together 'make up the evaluative framework through which we view the world' (A. M. Smith 2005, 251). These include not just our actively endorsed beliefs and other propositional attitudes but also our unconsciously held values and commitments which may only become apparent when we are faced with difficult decisions or demands to justify our actions. These high stakes scenarios, Smith claims, can reveal what we *really* are committed to: 'we may never have consciously entertained the judgments in question, and we may even be surprised by our own reactions in certain circumstances, precisely because they reveal rational or evaluative commitments which we were hitherto unaware of holding' (264).

On Smith's view, to be responsible for an action or attitude of hers, that action or attitude must reflect the agent's evaluative judgements. With regard to our attitudes, the notion of 'reflecting' is captured by the following kind of rational connection:

some of our mental states are linked to particular judgments in such a way that, if one sincerely holds a particular evaluative judgment, then the mental state in question should (or should not) occur. The 'should' in question here is the should of rationality and, therefore, marks a normative ideal which our actual attitudes may not always meet. (A. M. Smith 2005, 253)

That is to say, if the agent has certain values or commitments, in order to be rational she should acquire certain attitudes. We can say the same for actions: barring external

interference, if one holds certain evaluative judgements then, rationally, one should take certain actions and not others. Where the agent's behaviour reflects her evaluative judgements in this way, but those evaluative judgements are mistaken or otherwise problematic, the agent is culpable and open to criticism. Smith illustrates with a non-moral example: an agent who believes that spiders are dangerous may be perfectly rational in fearing spiders, but because her reason for that fear is a mistaken evaluative judgement, the fear itself is unwarranted, and she is open to criticism. For a moral example, a significant range of cases arises in connection with sexism and racism. The sexist or racist agent implicitly or explicitly holds mistaken judgements about women or racialized people, e.g., that they are unintelligent, dangerous, weak, and so on. Another way to capture the idea is that the agent has inadequate moral concern for some people (viz. women or racialized people). These judgements then rationally lead the agent to take discriminatory action or to acquire further objectionable attitudes. The rational connection between the sexist or racist agent's evaluative judgements and her actions or beliefs makes it appropriate to judge that she is responsible for these actions and attitudes, opens her to moral criticism for the same, and because these actions and attitudes are morally wrong, the agent is also culpable for them.

As is clear from the above, the rational relations view rejects the assumption central to voluntarism and the deep self view that culpability requires a voluntary choice to do something wrong or to accept aspects of one's psychology. At least with respect to the agent's attitudes and other mental states, the required rational relations may obtain between these states and the agent's evaluative judgements without any deliberate choice or reflective acceptance. Smith concludes that the rational relations view enables us to judge that the agent is responsible for 'ordinary cases of belief, intention, most desires, fear, indignation, admiration, and guilt, among others, as well as our moral perceptions and various patterns of unreflective thought and feeling which we take to be sensitive to and expressive of our underlying values and

commitments' (A. M. Smith 2005, 262–63). Culpable ignorance, on the rational relations view, is thus a particular case of responsibility for one's attitudes. Where the agent's lack of awareness that she is acting wrongfully is rationally connected to her evaluative judgements, she is culpable for her ignorance, and therefore culpable for the wrongful actions she takes from said ignorance. The form of this rational connection, Smith explains, is one of rational influence at an unconscious level: 'We assume that, if a person judges something to be important or significant, this should (rationally) have an influence on her unreflective patterns of thought and feeling concerning the thing in question' (255). For example, in *Peanuts*, Chuck is unaware that he is acting wrongfully in putting out peanuts at his party, and thus putting Lucille's life in danger, because he has forgotten that Lucille is allergic to peanuts. On Smith's view, Chuck is responsible for having forgotten about Lucille's allergy, because this failure to recall an important fact about his friend suggests that he does not, in fact, judge that his friends or their safety are particularly important. This error opens him to moral criticism for failing to remember Lucille's allergy, so he is culpable for forgetting her allergy, and thus culpable for putting her life in danger.

Smith's rational relations view also enables us to conclude that we are responsible for our conceptual incapacities. The agent's use of problematic concepts may directly reflect her mistaken evaluative judgements. The evaluative aspect of the concept *SAVAGE*, for example, is itself a negative valuation of indigenous people — a judgement that they should be discriminated against on the basis of stereotypical properties they are assumed to have, all by virtue of their social identity. The agent need not even be aware that she is making such an evaluation, for on Smith's view our evaluative judgements can operate below a conscious level. Smith's view also allows that the agent may be responsible for her conceptual non-competence or conceptual ignorance. If she is ignorant or non-competent because she has ignored or dismissed the testimony of hermeneutically dissenting groups regarding their experiences, for

instance, this may reflect an objectionable evaluative judgement on her part. This judgement may be that people in these groups often have strange or crazy social views, or that they are untrustworthy; or, the agent may simply lack adequate concern for people in these groups.

However, it is problematic that the rational relations view seems to require that in order to be responsible for something, that thing must, as Hieronymi puts it, '[reveal] what we find true or valuable or important' (Hieronymi 2014, 19). For if this is so, it means the rational relations view cannot accommodate a number of intuitive cases of culpability where the agent's behaviour does *not* reflect her evaluative judgements, but rather reflects a moment of *irrationality*. These cases are not those of *akrasia*, where the agent knows what is morally required of her but acts wrongly out of a conflicting desire. In these cases, her irrational action can be interpreted as reflecting *conflicted* evaluative judgements: on the one hand, she judges that she should behave one way because it is the right thing to do, on the other hand, she judges that the way she in fact behaves is preferable, and the wrong judgement ends up influencing her behaviour. But there are other kinds of irrational behaviour. Sometimes, we act wrongly despite having strong evaluative judgements against doing so, not because we have conflicting desires or judgements or anything else, but merely because, in this instance, our judgements fail to have the influence they should.

Recall the example of *Peanuts*. Above, I suggested that on Smith's view, Chuck may be culpable for forgetting Lucille's allergy, and thus for putting her life in danger, because his negligence may reflect some unrecognized lack of adequate moral concern for his friend or for safety. But this is not the only way to interpret the case. Chuck may care very deeply about Lucille and his friends generally, valuing their health and safety and usually being conscientious of their needs. On this alternative interpretation, he fails to remember something morally important (Lucille's allergy) not because he holds, 'deep down', some objectionable evaluative judgement with regard to Lucille or



towards his friends in general, but simply because he forgot. There is nothing mysterious about this. All of us forget important things from time to time, and this fact does not necessarily betray some malevolence or lack of concern that is hidden by our ordinary patterns of morally acceptable behaviour. Sometimes, we just screw up. Yet, Chuck and the rest of us still seem responsible for unwittingly doing harm when we forget such important things, even when this moment of forgetfulness is out of step with our judgements, values, or character. Perhaps where these moments *are* misaligned with what we take to be valuable or important, or with our previously demonstrated character, those affected will be more generous with their forgiveness — but forgiveness implies that there is a fault to be forgiven.<sup>75</sup>

Smith briefly touches on this sort of issue in her discussion of the kind of rational connection that, on the rational relations view, obtains between the agent's evaluative judgements and the behaviour for which the agent is responsible. Again, using the example of fear:

if I sincerely judge that there is nothing dangerous or threatening about spiders, I should not be fearful of them. The emotion of fear is conceptually linked to the judgment that the thing feared is in some way dangerous or threatening; therefore, my judgment that spiders are not in any way dangerous or threatening rationally entails that I should not be fearful of them. Most of the time this rational connection holds in fact: we are fearful of things we judge to be dangerous or threatening, and we are not fearful of things we judge not to be dangerous or threatening... Occasionally, however, this connection can fail: I may continue to fear spiders even though I judge them not to be dangerous or threatening in any way. When this happens, I am open to a particular kind of rational criticism — namely, to a charge of irrationality. (A. M. Smith 2005, 253)

Because Smith is committed to the view that attributability and answerability are aspects of the same phenomenon (cf. A. M. Smith 2012), she claims that being open to criticism for some action or attitude on the basis of irrationality is to be responsible for that irrational activity. But it is difficult to square this with her claim that when the

---

<sup>75</sup> Forgiveness where *no* fault has been committed can be a manipulative degeneration of this moral practice. See Fricker (2018).

agent is responsible for something, that thing must *reflect* or be *rationally connected* to the agent's evaluative judgements. In cases of irrationality, the connection appealed to seems to be a normative requirement that the agent has failed to meet. In finding a spider frightening despite judging that spiders are not dangerous, the agent's attitudes fail to be influenced by her evaluative judgements as they should. In forgetting Lucille's allergy, Chuck's mental states fail to be influenced by his evaluative judgements as they should. While it is fairly clear that this opens them to criticism, it is less clear that this constitutes a 'rational connection' between their evaluative judgements and their actions. What seems objectionable about their behaviour, rather, is that a rational connection *fails* to obtain between their judgements and their behaviour. And if the 'rational connection' in Smith's theory is just a normative requirement that the agent may or may not meet, then it is unclear that the rational relations view describes a condition for attributing responsibility. It seems rather to describe a norm of rationality that the agent has violated, for which the agent's responsibility remains an open question. Something further is required to explain what makes this failure of rationality attributable to the agent.

### 3.5. Psychologism

In response to the same worries raised earlier regarding culpable ignorance and voluntarism, Sher develops an alternative account of the epistemic condition on responsibility.<sup>76</sup> On Sher's view, the agent is morally responsible for some action when he fulfils any other conditions for responsibility and

- (1) is aware that the act is wrong or foolish when he performs it, or else,
- (2) is unaware that the act is wrong or foolish despite having evidence for its wrongness or foolishness his failure to recognize which
  - (a) falls below some applicable standard, and

---

<sup>76</sup> Sher does not defend a version of the control condition to accompany his epistemic condition. I discuss this in connection with responsibility for conceptual incapacities below, and in §3.6.

(b) is caused by the interaction of some combination of his constitutive attitudes, dispositions, and traits. (Sher 2009, 88)<sup>77</sup>

Sher's version of the epistemic condition supplements the voluntarist's — which is captured by (1) — with a condition for unwitting wrongdoing, which includes a normative condition (2a) and a causal condition (2b). Given the central importance of the causal structures constituted by the agent's psychological traits for attributing unwitting acts of wrongdoing to the agent, I will call Sher's account *psychologism*.<sup>78</sup>

The distinction made by (2a) and (2b) is that between the objective situation in which the agent finds herself (that which is 'external' to the agent), and subjective facts about the agent's moral-epistemic performance (that which is 'internal' to the agent). Among the things that are external to the agent, Sher includes not just the agent's surroundings, but also her physical traits, reasoning that 'few would deny that traits such as blindness and physical infirmity are external [to the agent] in a way that many psychological features are not' (Sher 2009, 105).<sup>79</sup> These external factors set the standards to which the agent is held: *any* agent in that situation, with the same evidence, would be morally required to be aware that taking that action is wrong. Among the things that are internal to the agent — things that are constitutive of her — Sher includes not just the agent's conscious experiences, explicit attitudes, and deliberations, but also any psychological traits that enable the agent to be consciously aware of what she is doing, and to respond to moral and other reasons for taking the

---

<sup>77</sup> Strictly speaking, this is only Sher's *partial* epistemic condition, concerning cases of morally bad behavior. His full epistemic condition includes a clause that allows for unwitting responsibility for morally *good* actions (q.v., Sher 2009, 143).

<sup>78</sup> This term is unfortunately misleading when taken out of context — by placing the emphasis on Sher's (2b) it obscures the role of the moral-epistemic norms of recognizing one's evidence captured by (2a) — but the term 'psychologism' still captures what is most distinctive about Sher's view.

<sup>79</sup> I do not agree with Sher that all physical features are obviously external to the agent's constitutive traits. Many view at least some of their physical features as partially constitutive of their self-identities — for example, see, among other works in disability studies, Rod Michalko's reflections on 'blindness as an essential part of being' (Michalko 1998, 4). But I will follow this assumption for the sake of argument.

actions that she does. This ‘enduring causal structure that *gives rise to* these responsibility-related activities’ (121) is what, on Sher’s view, ‘makes someone the particular responsible agent he is’ (133), and thus permits the attribution of the agent’s failure to recognize that he is acting wrongly *to him*. When these capacities are functioning optimally, the agent responds appropriately to her evidence for the rightness or wrongness of taking various actions, whether or not she is consciously aware that she is acting rightly or wrongly. But these traits do not always function optimally, and can cause failures to recognize that taking some action may be wrong. Where this failure falls below the moral-epistemic standards (which would apply to any agent in that situation) of what the agent should recognize, the agent is still responsible despite not being aware that she is acting wrongly. Only where the agent ‘simply lacks any information that would support the conclusion that he is acting wrongly’ (21), or where no agent in his circumstances could reasonably be expected to know that the act is wrong, is the agent exculpated on the basis of his ignorance.

Regarding *Peanuts*, psychologism delivers a more intuitively compelling judgement of Chuck’s culpability. Among the evidence Chuck has that putting out peanuts at his party would be wrong is Lucille’s past testimony that she has a serious allergy to them.<sup>80</sup> In Chuck’s rather ordinary conditions, we would expect any agent to remember such important information about his close friends — indeed, Chuck himself endorses this standard — even though what one remembers and forgets is not under one’s direct control. But Chuck has forgotten about Lucille’s allergy, and so is unaware that in putting out peanuts at his party he is acting wrongly by putting his friend’s life in danger. His forgetting is caused by the interaction of his constitutive psychological traits — perhaps including his belief that he would remember anything important like

---

<sup>80</sup> This depends a bit on how we construe the relevant notion of ‘evidence that is available to the agent’. But I do not think it is strange to include things that the agent has heard and should remember.

this, his long-term memory, and any unconscious cognitive processing that should have signaled (but failed to signal) this morally important fact to his conscious deliberation. So despite being unaware that he was acting wrongly in putting out peanuts at his party, Chuck is still morally responsible for doing so, and thus for putting Lucille's life in danger. The act of wrongdoing can be attributed to his agency because, according to psychologism, these conscious and unconscious psychological attitudes, dispositions, and traits are all (partially) constitutive of who he is.

Moreover, whether or not we suppose that Chuck's forgetfulness reflects some underlying evaluative judgement (either about Lucille or about safety), we can still find Chuck responsible for putting Lucille's life in danger. In the case where Chuck has acted rationally, say because he does not attach adequate moral value to Lucille's safety, this judgement may be among the psychological causes of his forgetfulness. In the case where Chuck has acted irrationally, because he holds no problematic evaluative judgements, it suffices to point to his long-term memory as one of the psychological causes of his forgetfulness. Either way, the causal condition (2b) is fulfilled, and since we would expect anyone in a similar situation to remember Lucille's allergy, the normative condition (2a) is also fulfilled. So Chuck is culpable for not being aware that he is acting wrongly, and thus for putting Lucille's life in danger.

Psychologism looks more promising as an account of our responsibility for our conceptual incapacities. Of course, it is not a complete theory; Sher does not defend a version of the control condition on responsibility. Rather, he briefly argues that his view is not incompatible with various contenders for the control condition (Sher 2009, 145–52). But if we can interpret (2) in his epistemic condition as giving us the conditions for culpable ignorance, we can use this theory to accommodate the intuition that we are sometimes responsible for our mental states, despite not usually having

control over them.<sup>81</sup> With this in mind, we can turn to responsibility for our conceptual incapacities.

Suppose the agent lives in a liberal democracy and is not somehow evidentially isolated (e.g., because of living in a commune cut off from wider society). Given her moral-epistemic environment, she has ample evidence that using a concept like *SAVAGE* would be problematic; she has encountered plenty of evidence that indigenous people are not sub-human. But she has failed to pick up on this fact and continues to use the wrongful concept. This moral-epistemic failure clearly falls below the standards which apply to her: the evidence is enough that she should have realized that using this concept is wrong. Her failure to recognize this is also due to some combination of her psychological traits. These may include prejudices that she holds, or she may have simply failed to make the connection. In any case, her ignorance is not due to mind-control or some other interference that bypasses her psychological traits. So, according to psychologism, she is responsible for using this concept. Similarly, we may show that she is culpable for her conceptual ignorance or non-competence. Her lack of awareness that she should use certain concepts, or that she holds various mistaken beliefs about a concept, may also be caused by her psychological traits and fall below an applicable standard given the evidence available to her.

Psychologism runs into trouble, however, in accommodating our intuitions regarding unwitting wrongdoing in cases of agents with mental disorders.<sup>82</sup> When the

---

<sup>81</sup> The same intuition motivates the rational relations view; see Smith (2005).

<sup>82</sup> I owe the observations that Sher's view requires us to view agents with mental disorders as culpable for their ignorant wrongdoing, and that activists in the neurodiversity movement view these conditions as among those constitutive of their identities, to Mathieu Doucet and Dylan McChesney's talk at the 2017 Congress of the Canadian Philosophical Association at Ryerson University. I have benefitted greatly in this section from reading a subsequent draft they presented at Sheffield in May 2018 (Doucet and McChesney, ms., cited with permission). While my worries about Sher's view are broadly the same as theirs, my positive proposal (§3.7) is substantively different: they opt for a 'moral concern' view of responsibility that is inspired by Angela Smith and Nomy Arpaly. I thank Matt Doucet for additional discussion.

agent's wrongdoing — ignorant or otherwise — is caused by a mental disorder of hers, we generally take this to be an excusing factor. This is relevant to our inquiry into responsibility for conceptual incapacities when we recognize that some mental disorders may cause us to use certain concepts in our thinking. For example, some mental disorders produce thoughts that the agent finds intrusive and upsetting — and even neurotypical agents experience such intrusions from time to time. One example is the so-called 'high place phenomenon': while standing at the edge of a precipice (or when similarly confronted with a dangerous situation) one may suddenly think of jumping off (or otherwise doing something suicidal). Such thoughts have been reported both among people experiencing regular suicidal ideation and the general population (Hames et al. 2012). Similarly, thoughts constructed using problematic concepts could spontaneously occur to the agent because of a mental disorder. But, intuitively, agents are not culpable for actions or attitudes of theirs which were caused by mental disorders. So we should aim for an account that does not find agents with mental disorders culpable for their intrusive thoughts which contain problematic concepts.

Unfortunately, psychologism gives us the wrong answer. Consider the following variation on *Peanuts*:

*You're OCD, Charlie Brown.* Charles is organizing a party at his home. One of his invited guests is his childhood friend Lucille, who is seriously allergic to peanuts. Lucille has told Charles about her allergy several times in the past, but it has slipped his mind — it has been some time since she has explicitly mentioned it, but she quite reasonably assumes that good ol' Charles will remember. He does not. Charles has an undiagnosed case of obsessive-compulsive disorder, which manifests in a cascade of obsessive thoughts about whether his friends will show up, whether they will have a good time, whether they actually like him, whether his home is clean enough, whether the decorations are sufficient, and so on. While he believes that he will and should remember morally important facts such as his friends' dietary requirements, his OCD overtakes his attention and unconscious cognition, preventing him from remembering what Lucille has told him. The snacks he leaves out at the party include peanuts, Lucille comes into contact with them, and she goes into anaphylaxis.

Charles's OCD seems intuitively to provide him with an excuse. What we want to say is that it is not his fault he was unable to recall Lucille's allergy: his OCD interfered with his ability to remember, and conditions like OCD are out of the agent's control. But this formulation of the excuse is not permitted on Sher's account. Charles, like Chuck, has evidence that putting out peanuts where Lucille could come into contact with them would be wrong, the applicable standard is the same (we would expect anyone in Charles's situation to recall that Lucille has a serious allergy to peanuts), and Charles's ignorance is caused by the interaction of his psychological traits, *among which is his OCD*. The fact that Charles's OCD and what it makes him able to remember are out of his control is irrelevant to psychologism. On this view, he is culpable for his ignorance and thus for putting Lucille's life in danger, *despite* his ignorance being caused by a mental disorder.

Sher seems willing to bite the bullet with regard to cases like *You're OCD*, *Charlie Brown*, though he also suggests a reply to the worry that we do not in fact consider such agents responsible for their unwitting wrongdoing. In one of his own examples, the agent who unwittingly acts wrongly could be interpreted as doing so out of a mental disorder:

*Home for the Holidays*. Joliet, who is afraid of burglars, is alone in the house. Panicked by sounds of movement in her kitchen, she grabs her husband's gun, tiptoes down the stairs, and shoots the intruder. It is her son, who has come home early for the holidays. (Sher 2009, 26)

Joliet's behaviour could potentially be explained by one of several mental disorders, such as paranoid personality disorder or a some form of anxiety disorder, all of which may cause behaviour and beliefs that are motivated by a fear that is out of proportion with the actual danger or risks (American Psychiatric Association 2013). Sher is willing to accept that Joliet is responsible for shooting her son, but suggests that, if he is wrong about our intuitions regarding this and similar cases, we may excuse agents whose unwitting wrongdoing is caused by their mental disorders by drawing a line between



the psychological traits that are constitutive of the agent and those that are external to the agent. Sher writes that our hesitation to judge that agents like Joliet and Charles are responsible for their unwitting wrongdoing

is naturally attributed to our uncertainty about the boundaries of the self. Because we are not sure on which side of that boundary to locate the traits that explain why Joliet [and Charles] have not recognized their acts or omissions as wrong, the account correctly predicts that we will lack confidence in our judgements that they are responsible *for* their wrongful acts or omissions. (93)

Despite the uncertainty as to whether mental disorders are partly constitutive of the agent or external to her, if we assume that they are external, then they become part of the conditions we factor into the objective situation by which we set the standards for recognizing morally relevant facts. Agents with mental disorders would therefore be subject to different applicable standards, such that we would no longer expect them to be aware that they are acting wrongly. Because Charles's OCD sometimes makes it more difficult for him to remember morally important facts, he is not subject to the same standards as the neurotypical Chuck, who has no such difficulties and thus is expected to remember such things.

However, it is not clear that this move is defensible with regard to mental disorders. First, there are principled reasons we might want not to exclude these conditions from the psychological traits that are constitutive of the agent, while still holding on to the intuition that they excuse the agent for unwittingly doing something wrong. Activists in the neurodiversity and Mad Pride movements, for instance, argue that gaining wider acceptance of people with various conditions that have been classified as mental illnesses or disorders involves acceptance of these conditions as part of what makes these people who they are. As autistic writer John Elder Robison writes,

To many neurodiversity proponents, talk of 'cure' feels like an attack on *their very being*. They detest those words for the same reason other groups detest talk of 'curing gayness' or 'passing for white', and they perceive the accommodation of neurological differences as a similarly charged civil rights

issue. If *their diversity is part of their makeup* they believe it's their right to be accepted and supported 'as-is'. They should not be *made into something else* — especially against their will — to fit some imagined societal ideal. (Robison 2013, my emphases)

The emphasized phrases indicate that neurodiversity advocates view mental illnesses and disorders as constitutive parts of their identities, in a similar capacity to other social group identities. What is distinctive about mental disorders, and what these activists are careful not to deny, is that these conditions can lead to behaviour and patterns of awareness that are morally problematic. By that token, people with mental disorders may require some support to prevent any harmful manifestations of these conditions, but, neurodiversity advocates argue, *without* the typical presumption in favour of eliminating the condition. Whether or not we agree with the mission of the neurodiversity movement in general or with the application of its principles to the particular mental disorders discussed above, it is clear that for Robison and others dedicated to improving conditions for neurodivergent people, these conditions *are* part of the constitutive traits of these agents. Indeed, for many of these activists, these conditions are an overlooked source of value and the basis of communities of solidarity.<sup>83</sup> Yet at the same time, the support and understanding needed by those with conditions like these, and the pointlessness of blame for these behaviour caused by them — a mental disorder is not responsive to criticism the way we expect false beliefs to be — strongly speaks against the judgement that people with mental disorders, such as Charles, are culpable for their unwittingly wrongdoing.

Second, it is not clear how Sher's view can demarcate which psychological traits are constitutive of the agent and which are not, such that mental disorders are excluded. Sher raises the possibility that some psychological features may 'have no obvious connection to any of [the agent's] beliefs about himself or the world, his

---

<sup>83</sup> For another example, see the Toronto Mad Pride organization's mission statement, in which this conception of one's mental illnesses and disorders as partially constitutive of one's identity is implicit throughout (Toronto Mad Pride 2017).

judgements about what he has reason to believe or do, or his actual decisions, actions, or omissions', and thus may be excluded from those psychological traits that are relevant to the agent's moral responsibility for his actions, omissions, and beliefs (Sher 2009, 119). But the problem is precisely that mental disorders such as OCD, anxiety disorder, and paranoid personality disorder *do* have obvious connections to the agent's beliefs, judgements, and actions. These conditions and many other mental disorders interfere with the agent's judgements of what to believe and what to do, as we saw with Charles and Joliet. Psychologism does not provide the resources to differentiate mental disorders from other psychological traits in a way that makes a difference to attributions of responsibility. As a consequence of this view, there is no meaningful difference, from the perspective of responsibility-attribution, between a mental disorder and a moral vice. Both are part of the agent's constitutive psychological traits that causally produce her patterns of conscious and unconscious recognition of morally important facts. One kind of vice is a stable disposition to come to false beliefs about what one should do and to take those actions, in spite of the available evidence — and some mental disorders seem to produce the same outcome.

A defender of psychologism might try to accommodate the difficulties posed by mental disorders by adding the possibility that some factors can *mitigate* the agent's culpability, without actually excusing her from being responsible. Or we might bite the bullet and admit that the agent with a mental disorder is responsible, but *deny that she is at fault*. Or, we might introduce some way of recognizing *excuses* other than simply lacking evidence that one is acting wrongly, such that the agent would be excused for her actions or attitudes caused by her mental disorders. Sher's own account does not include any mechanism for recognizing degrees of culpability, responsibility without fault, or excuses other than lacking evidence, but his view is not obviously incompatible with these additions. The strategy I take, in the next section, is to introduce a way to excuse behaviour caused by a mental illness.

### 3.6. A Hybrid View: Reasonable Psychologism

So far we have rejected voluntarism and the deep self view on the basis that, among other problems, they fail to give an adequate account of culpable ignorance. Although the rational relations view delivers a more satisfying account of culpable ignorance, we have seen that it encounters difficulties explaining culpability for actions done irrationally, which do not reflect the agent's evaluative judgements. Psychologism fares better than the rational relations view, with respect to both culpable ignorance and irrational behaviour, but it is too weak as a theory of the epistemic condition because it leads to a dilemma: either agents with mental disorders are culpable for their unwitting wrongdoing caused by their mental disorders, or those mental disorders are not part of the psychological traits that make them who they are. In this section, I sketch my proposed solution: combine elements of the rational relations view with psychologism to overcome the shortcomings of each. I call this view *reasonable psychologism*. Fully defending this view is not possible in the present context; for my purposes, it will suffice to have a defensible outline.

First, to get an idea of how we might strengthen psychologism in the appropriate way, let's look at how the rational relations view handles cases of agents with mental disorders, such as Charles. On the one hand, it is possible that these agents are excused from being responsible because those of their actions that are caused by their mental disorders do not reflect their evaluative judgements. Charles, perhaps, has a perfectly adequate level of moral concern for Lucille, but his OCD affects his ability to remember her allergy. So while one of his constitutive psychological traits caused him to forget something he should have remembered, because his forgetfulness does not reflect his evaluative judgements, he is not responsible.<sup>84</sup> But as we saw in §3.4, the

---

<sup>84</sup> This appeal to the mentally ill agent's evaluative judgements is the strategy adopted by Doucet and McChesney (ms.) for avoiding the problems posed by psychologism. They find it an acceptable consequence of rational relations view that agents with mental disorders are morally responsible for actions of theirs that *do* reflect their evaluative judgements, even where those judgements are caused by their mental

rational relations view's reliance on this sort of connection between the agent's behaviour and her evaluative judgements is too strong. For, requiring that there be such a rational connection fails to capture cases of responsibility where the agent has been irrational. Moreover, the rational relations view would require us to conclude that agents with mental disorders are responsible when their actions reflect evaluative judgements they have acquired *because of their mental disorders*. Some mental disorders exert such influence: an arachnophobe may judge spiders to be dangerous precisely because of her condition. And Charles could be interpreted as judging, however implicitly, that all the details of his party over which he frets are more important than Lucille's safety, precisely because of his OCD.

Smith briefly notes this last worry for her view, but does not develop a response in any detail: 'The attitudes of the OCD agent... appear to reflect her own (albeit mistaken) evaluative judgments. Does it not follow that such an agent is [responsible] for such attitudes on my view?' (A. M. Smith 2012, 585 fn. 13). She suggests that what excuses the agent with a mental disorder is the fact that she is not *answerable* for these judgements, because she is not answerable for the condition that gives rise to them (585). We cannot make this move on my account, however; recall that in §3.1 I argued that responsibility attribution is a distinct notion from that of having obligations arising out of one's answerability. Still, Smith's invocation of answerability, in the sense that the agent is open to demands for her *reasons* for doing something, is suggestive. Elsewhere, Smith argues that if the agent does something bad in a moment of 'severe mental delusion', we would not view her as responsible for this behaviour because, in this moment, the agent was not 'reasons-responsive' (A. M. Smith 2008, 388). This suggests that a more fundamental condition for being responsible for

---

disorders. On their account, this contributes to the neurodiversity ambition of recognizing agents with mental disorders as fully responsible agents. But I do not think this worry applies to reasonable psychologism, as I develop it. See fn. 86, below.

something is that the agent was, at the time of her doing thus-and-so or acquiring such-and-such an attitude, *responsive to reasons*.

What does it mean for some attitude, disposition, trait, or other element of the agent's psychology to be 'reasons-responsive'? This is a complex and subtle issue that Smith does not elaborate upon. One approach would be to draw on the account developed by John Fischer and Mark Ravizza (Fischer and Ravizza 1998). On their view, reasons-responsiveness has two components: reasons-receptivity and reasons-reactivity. The former concerns a psychological mechanism's influence over the agent's recognition of her reasons for acting: a mechanism is *reasons-receptive* where, in a sufficient range of counterfactual scenarios where that mechanism operates, the agent recognizes her reasons for (or against) taking some action or acquiring some attitude. The latter concerns a psychological mechanism's influence over the agent's choice to act in accordance with her reasons: a mechanism is *reasons-reactive* where, in a sufficient range of counterfactual scenarios where that mechanism operates, the agent chooses to act upon the reasons she recognizes. Mental disorders may fail to be reasons-receptive — they prevent the agent from recognizing her reasons — or they may fail to be reasons-reactive — while the agent recognizes her reasons, they prevent her from acting upon those reasons.

Fischer and Ravizza's emphasis on *recognition* and *choice* in their notion of reasons-responsiveness, however, is incompatible with the move away from voluntarism that motivates both psychologism and the rational relations view. It makes little sense to say that Charles is not responsible for forgetting about Lucille's allergy because his OCD made it such that he could not *choose* to remember that she is allergic to peanuts. This aspect of Fischer and Ravizza's account is a reflection of their primary interest in responsibility for actions, rather than attitudes or other mental activities. But we can fix these problems by modifying the idea of reasons-responsiveness in the light of elements that are common to both the rational relations view and

psychologism. First, in order to include the idea that we can be responsible for our mental activities as well as our actions, we can collapse reasons-receptivity and reasons-reactivity. Recognizing one's reasons is just a form of reacting to one's reasons at the level of mental activity. Second, we can add that many of the psychological mechanisms by which we respond to reasons in our actions and mental activities operate unconsciously. Responding to one's reasons in one's actions or attitudes can happen without being aware of the reasons to which one is responding. When we add these together, we have: a psychological mechanism is *reasons-responsive* when, in a sufficient range of counterfactual scenarios in which that mechanism operates and where the agent has reasons to  $\Phi$ , the agent  $\Phi$ s.

Incorporating this notion of reasons-responsiveness, we can modify the epistemic condition defined by psychologism as follows:

*RP-E*: An agent *A* is responsible for some action when *A* meets all the non-epistemic conditions for responsibility and:

- (E1) *A* is aware at the time that the action is wrong; or
- (E2) *A* is unaware that the action is wrong, despite having evidence for its wrongness, her failure to recognize which
  - (E2a) falls below an applicable standard, and
  - (E2b) is caused by some combination of *A*'s *reasons-responsive* constitutive attitudes, dispositions, or traits.

By restricting the range of psychological mechanisms referred to in psychologism's (2b) to those psychological mechanisms which are reasons-responsive, *RP-E* acknowledges that mental disorders are part of the agent's constitutive traits while at the same time denying that the agent is responsible for activities of hers that are caused by these conditions. This is because mental disorders are not reasons-responsive. They can prevent the agent from being aware that she is acting wrongly, as when Charles's OCD causes him to fail to remember that Lucille is allergic to peanuts. *RP-E* also delivers intuitive answers to other cases. Chuck is responsible for putting out peanuts at his party and thus for putting Lucille's life in danger, because his failure to

recognize that he was doing something wrong is due to a failure of his memory, a reasons-responsive psychological mechanism.

But *RP-E* is not quite enough. For one thing, we are still working with responsibility for actions, and we need an account that acknowledges that we can be responsible for our mental activities as well. For another, it may be the case that in many counterfactual scenarios where the agent's mental disorders operate, the agent *is aware* that she is acting wrongly, but is unable to act upon this recognition of her reasons not to take that action. For example, in a modified version of *You're OCD, Charlie Brown*, Charles may remember that Lucille is allergic to peanuts, but be unable to act upon this information because his OCD interferes. It is less clear in this variant case whether Charles is excused from responsibility, because it is unclear whether his OCD makes it *impossible* to act on his reasons, or merely *difficult* to do so. These two issues seem to pull in opposite directions. On the one hand, incorporating a means to excuse agents whose mental disorders make it impossible for them to act on reasons they recognize seems to call for a control condition. On the other hand, acknowledging responsibility for attitudes and other mental activities seems to require denying that control is necessary for responsibility, at least in the case of mental activity.

But these ambitions are not in tension. We can incorporate both by using the notion of reasons-responsiveness defined above to weaken the rational relations view and bring it closer to the modified version of psychologism expressed by *RP-E*. Recall that the rational relations view appeals to rational connections between the agent's evaluative judgements and the agent's activities as the condition for responsibility attributions. We can expand this notion of a 'rational connection' to include not just the influence of the agent's evaluative judgements, but also the causal influence of any of the agent's reasons-responsive attitudes, dispositions, traits, or other psychological mechanisms. So we have:



*RP-C*: An agent *A* is responsible for  $\Phi$  when  $\Phi$  was caused by some combination of *A*'s reasons-responsive constitutive attitudes, dispositions, or traits.

*RP-C* includes but also moves beyond the notion of conscious control. According to *RP-C*, *A* would be responsible for an intentional action because such an action is caused by whatever combination of psychological mechanisms constitutes intentionally taking an action. *A* would not be responsible where her action was under coercion, however, because the cause of her intentional action in that case is the coercive pressure. *A* would also be responsible for attitudes she acquires, for her patterns of awareness, and for any other mental activities, where these are caused by her reasons-responsive psychological mechanisms, including her memory, perception, reasoning, and so on.<sup>85</sup>

We can now combine *RP-C* and *RP-E* into a theory of responsibility attribution as follows:

*RP*: An agent *A* is responsible for  $\Phi$  when:

(C)  $\Phi$  is caused by some combination of *A*'s reasons-responsive constitutive attitudes, dispositions, or traits; and

(E1) *A* is aware that  $\Phi$  is wrong at the time, or

(E2) *A* is unaware that  $\Phi$  is wrong, despite having evidence of  $\Phi$ 's wrongness, her failure to recognize which

(E2a) falls below an applicable standard and

(E2b) is caused by some combination of *A*'s reasons-responsive constitutive attitudes, dispositions, or traits.

This theory delivers the correct answers in all the standard and problem cases discussed above. Suppose the agent knowingly acts wrongfully; for example, imagine that in the scenario presented by *Peanuts*, Chuck is entirely innocent but the devilish Charlotte intentionally sabotages his party by putting out peanuts, knowing full well, indeed hoping, that this may cause Lucille to have an allergic reaction, and knowing that it is wrong to do so, but not caring, so much is her desire to ruin Chuck's party. This action is caused by those of Charlotte's reasons-responsive constitutive psychological traits that are involved in intentional action. Moreover, she is aware that

---

<sup>85</sup> Cf. Conor McHugh's theory of doxastic responsibility, which he bases on Fischer and Ravizza's account of reasons-responsiveness (McHugh 2013).

she is acting wrongly. So Charlotte meets conditions (C) and (E1), and is therefore responsible. In the unmodified version of *Peanuts*, Chuck puts Lucille's life in danger despite not being aware that he is doing so. His putting out peanuts at his party was caused by his reasons-responsive constitutive psychological traits. His lack of awareness that doing so was wrong, despite having evidence of this, falls below the applicable moral-epistemic standards and is caused by his reasons-responsive constitutive psychological traits. So he meets conditions (C) and (E2), and is therefore responsible. This holds whether or not his action is radically out of character. In a modification of *Peanuts* where neither Chuck nor Lucille know that Lucille is allergic to peanuts, Chuck is excused. This is because he simply lacks evidence that putting out peanuts at his party would be wrong, and therefore fails conditions (E1) and (E2). In *You're OCD, Charlie Brown*, Charles is excused. This is because his failure to remember that Lucille is allergic to peanuts is caused by a constitutive psychological trait of his that is *not* reasons-responsive, namely, his OCD. He therefore fails condition (E2b). In the modification of this case where Charles remembers that Lucille is allergic to peanuts, but is unable, because of his OCD, to act upon this recognition that putting out peanuts would be wrong, the cause of his action is a constitutive psychological trait of his that is *not* reasons-responsive. He is thus not responsible because he fails condition (C). Where his OCD merely makes it difficult for him to act on his awareness of Lucille's allergy, he would meet conditions (C) and (E1), making him responsible — though the difficulty he faces may incline us to blame him less or to forgive him more readily.<sup>86</sup>

A slightly trickier range of cases is presented by agents who act out of vices. For these attitudes, dispositions, and traits may seem to *prevent* the agent from responding

---

<sup>86</sup> This consequence shows that according to reasonable psychologism, there are cases where agents with mental disorders are not excused on the basis of their condition. This is in line with the neurodiversity commitment stressed by Doucet and McChesney (ms.), that an account of responsibility attribution should not lead to the conclusion that agents with mental disorders lack moral agency, while at the same time acknowledging that mental disorders sometimes exculpate.

to her reasons across most counterfactual situations in which they operate. In that case, does that not mean that these mechanisms are not reasons-responsive, thereby excusing vicious agents? We can get around this difficulty by noting that these psychological mechanisms are reasons-responsive, but *misdirected* — the reasons they respond to are insufficient to *justify* the agent's activity. For example, suppose Charline, because she holds a racist vice, judges that Lucius, a black man, is untrustworthy. Charline has reasons for this judgement, to which her psychological traits are responding, but they fail to justify it because they are mistaken and immoral assumptions about the character of black people. As Fischer and Ravizza argue, so long as the agent's responsiveness forms a 'minimally comprehensible pattern' that is not 'utterly divorced from reality', we can count her as responding to reasons, even if they are bad reasons (Fischer and Ravizza 1998, 73). This observation also allows reasonable psychologism to work around cases where the agent acts on 'weird' reasons.<sup>87</sup> For example, imagine an agent who takes the fact that a woman in a red dress left a Tesco flyer in his mailbox to be a reason to destroy his neighbour's magnolias. Barring some further information that could make sense of this behaviour (perhaps he and the flyer-deliverer are secret agents who have agreed that her delivering a Tesco flyer while wearing a red dress is a code indicating that counterintelligence has bugged the magnolias), this sort of bizarre activity is difficult to comprehend and seems not to be grounded in reality, even if it can be stated in terms of responding to 'reasons'.<sup>88</sup>

---

<sup>87</sup> Sher offers an alternative solution to dealing with weird reasons, arguing that so long as the agent is, in principle, capable of making *correct* judgements about what she should do or what counts as a reason, she is still responsible (2009, 131–33). Cf. Wolf's 'sanity' requirement for responsibility (Wolf 1988).

<sup>88</sup> There is a potential worry here in connection with hermeneutical injustice. It is possible that sometimes the agent's reasons will not be intelligible to us because, as a result of her hermeneutical marginalization, the interpretive tools we need to make sense of her reasons are not available to us. But this worry is mistaken. What is required is that the agent's reasons are intelligible *in principle*. If we do not find her reasons intelligible because we lack further information due to hermeneutical injustice,

Finally, this mention of the agent's justificatory reasons brings us to reasonable psychologism's connection to answerability and accountability. Recall that the initial inspiration for adding reasons-responsiveness to psychologism was Smith's observation that an agent with a mental disorder does not seem open to demands that she justify her behaviour by providing her reasons. I suggest that when the agent is responsible for one of her activities (in the attributability sense captured by *RP*), the *reasons-responsiveness* of the psychological mechanisms of the agent — which caused that activity or her (lack of) awareness that she was doing wrong — is what opens her to questions of the form, 'Why did you do that?' or 'Why do you believe that?' That reasons-responsiveness puts her in a position where she *should* be able to provide a justificatory reason for her actions, attitudes, or other activities. Where she cannot provide reasons that justify her activity — because she did it for no reason or because she did something despite having reasons not to, whether or not she was aware of this — she is open to criticism. Along with criticism may come reproach, anger, or other blame responses which are aimed at holding the agent accountable.

The view I have sketched here under the label of reasonable psychologism is, of course, incomplete. Much more could be said about the notions of reasons, causes, reasons-responsiveness, and justification than I have adumbrated. But it suffices for present purposes to have a workable outline of a theory of responsibility attribution that avoids obvious problems, i.e., dramatically counter-intuitive results. Let us turn, then, to the issue of responsibility for our conceptual incapacities, according to reasonable psychologism.

---

then just as in the case of the secret agents, the agent's reasons turn out not to be disconnected from reality after all.

### 3.7. Responsibility for Conceptual Incapacities

The clearest cases of culpability for one's conceptual incapacities are those where the agent deliberately decides to use a particular concept. For example, imagine that a judge is deciding how to rule in a sexual assault trial. Part of the prosecution's strategy is to argue that the defendant routinely sexually harassed the complainant in the months before the alleged assault, thereby showing that he exhibited a pattern of sexual misconduct and strengthening their case with regard to the alleged assault. The defence counters that their client was merely flirting in those prior instances, and never acted with impropriety. Suppose that some aspect of the judge's overall ruling (perhaps whether to convict, or how harsh a sentence to impose) will be influenced by whether the judge sides with the prosecution's or the defence's interpretation of these events leading up to the alleged assault. That is to say, the judge's decision as to which concept applies to the defendant's prior behaviour, *SEXUAL HARASSMENT* or *FLIRTATION*, will affect the outcome of the case. Let's say that the prosecution's case provides sufficient evidence such that the judge should conclude that the defendant's actions constituted sexual assault. But when judge deliberates, he decides wrongly by siding with the defence, delivering an inappropriate ruling. Even supposing he is *not* non-competent with regard to the concept *SEXUAL HARASSMENT* — he simply makes a mistake — on the view I have defended, the judge is responsible for this error. The judge's decision to interpret the defendant's behaviour as flirtation is caused by his deliberative processes, which are reasons-responsive, so he meets condition (C). He also meets condition (E2). The judge has evidence that interpreting the defendant's behaviour as flirtation is wrong, namely, the prosecution's case. His lack of awareness that he has made the wrong interpretation falls below an applicable standard, for the prosecution's case is strong, he has put in sufficient time to his deliberation to consider all aspects of the case, and he has professional obligations to side with the prosecution when they prove their case beyond a reasonable doubt. The judge's deliberative

processes, which cause him to fail to realize that his interpretation is wrong, are reasons-responsive — in a wide range of counterfactual conditions where he deliberates on this case, he indeed sides with the prosecution — in the actual scenario he makes a mistake and does not react to those reasons. Since he meets conditions (C) and (E2), according to *RP* he is responsible for deciding to use the wrong concept.

Though cases of deliberately choosing to use a particular concept are the most clear cases of culpability, they are far from the most common. Usually, as is the case in the acquisition of most of our beliefs and attitudes, our cognition deploys our concepts without conscious intervention. A racist, for example, may deploy the concept *SAVAGE* in her attitudes about indigenous people without deliberating about whether to use a different concept. While not the result of a deliberate choice, this use of a wrongful concept is still something for which the racist agent is responsible. Because her use of this concept is caused by some of her reasons-responsive constitutive attitudes — namely, her belief that indigenous people are inferior and deserving of discriminatory treatment because of their social identity, which corresponds to the evaluative aspect of the concept — she meets condition (C). She also meets condition (E2). Assuming she lives in a social context largely like ours, she has plenty of evidence that should lead her to recognize that it is wrong to think of indigenous people as ‘savages’ — her failure to recognize this falls below an applicable standard. Her ignorance is also caused by some of her reasons-responsive constitutive traits, namely, those attitudes and dispositions that constitute her racism. So, according to *RP*, she is responsible for her use of the concept *SAVAGE*.

It is also possible that an agent who holds no explicit beliefs disparaging indigenous people may employ the concept of a ‘savage’ without realizing it. As Mark Cronlund Anderson and Carmen Robertson show in their study of the portrayal of First Nations people in Canadian newspapers, the image of the ‘savage’ continues to exercise a strong influence on the ways indigenous people in Canada are represented in the

media. The pejorative label 'savage' rarely appears in recent articles, but the same stereotypes of indigenous people as violent, promiscuous, biologically inferior, prone to alcoholism, and so on, continue to be raised indirectly:

The vivid sketching of Natives as savages... has transmogrified from a time when the news language left negligible room for ambiguity, to a situation in the early twenty-first century in which it may no longer be considered appropriate to apply such candid, distasteful monikers... the savage becomes one who engages in behaviour that defines savagery but is not referred to openly as a 'savage' ... it matters little whether you are called a savage or merely described as behaving like one... the mere association of Natives and, say, alcohol in a news story dredges up the drunken Native stereotype as surely as the beating of tom-toms announces an impending 'Injun' attack in an old Western. (M. C. Anderson and Robertson 2011, 269–70)

Anderson and Robertson analyse news consumers' interpretations of these articles' oblique references to 'savagery' by analogy to genre conventions. Just as a film might use a combination of props, scenery, and character tropes to signal that it is a Western (revolvers, a saloon in a dusty town, a swaggering gun-slinging young white man), a news article signals that the indigenous people it refers to are 'savages' by associating them with drugs or alcohol, crime-ridden neighbourhoods, mystical ancient traditions, violence at land-claims protests, and so on. On my account, we can interpret these portrayals of indigenous people as covertly using the thick concept SAVAGE, even though the text and possibly even its writers' and readers' thoughts do not explicitly employ the term 'savage'. The negatively valuated stereotypes that partially constitute that concept have persisted largely unchanged despite shifts in the language used to conjure them. Readers who have internalized Canadian culture's associations between indigenous people and these negatively charged stereotypes may be responsible for employing this wrongful concept even though they may not realize that their thinking has been driven in this direction. In reading such texts, these readers' reasons-responsive constitutive attitudes and dispositions pick up on these stereotypes and their implicit valuation, causing her to employ the concept SAVAGE — which may happen even if the agent explicitly disavows racist beliefs. Despite possibly not being

aware that she is doing wrong in using this concept, the Canadian reader will have a preponderance of evidence that thinking of indigenous people as ‘savages’ is wrongful, so her ignorance — caused by some of the same reasons-responsive constitutive traits that cause her to use the concept — is not an excuse. She thus meets conditions (C) and (E2), and is therefore responsible for using the concept *SAVAGE*.

What about cases where the agent’s intrusive thoughts deploy problematic concepts? Suppose the agent has a dissociative disorder, the key symptoms of which include mental and behavioural intrusions and a loss of control over mental functions (American Psychiatric Association 2013). If this disorder causes the agent to have thoughts constructed using a wrongful concept such as *SAVAGE*, she would be excused because her disorder is not reasons-responsive, so she fails condition (C). A neurotypical agent may also be excused for occasional intrusive thoughts using such concepts. It is not unusual for our minds to throw up odd, even unpleasant thoughts and images from time to time, without any reason. Where such thoughts genuinely are mere mental noise, they would not be caused by any reasons-responsive psychological process, so the agent would fail condition (C). But we should be careful here, because it is often difficult to tell whether these thoughts are genuine intrusions or reflect some pernicious psychological mechanism that is reasons-responsive but misdirected, such as an implicit bias or prejudice. We should be vigilant to patterns in these thoughts, which may indicate the presence of such processes.

Turning next to responsibility for conceptual ignorance and conceptual non-competence, recall the kinds of cases raised in the previous chapter in connection with hermeneutical injustice. Because of the hermeneutical marginalization of certain social groups, the concepts and conceptions that they have developed to interpret their distinctive and important social experiences may not be widely shared outside of their own communities. Pohlhaus and Dotson each claim that there are numerous cases where agents in relatively dominant social locations may be culpable for their



ignorance or non-competence with regard to these marginal concepts (Pohlhaus 2012; Dotson 2012). As we saw in Chapter 2, voluntarism makes culpability in these cases vanishingly rare. But on the view of responsibility I have defended, culpability for conceptual ignorance and non-competence is much more common.

Some agents will be conceptually ignorant or non-competent because of a deliberate choice to resist taking up marginal conceptual resources. This is not usually an instance of intentional wrongdoing, since these agents typically present a rationale for their resistance. Some conservatives, for example, reject LGBTQ+ concepts because the categories and experiences that these concepts identify are inadmissible on traditional conceptions of gender. For some, this resistance maintains ignorance of those marginal concepts and conceptions such that they simply do not know what concepts such as TRANSGENDER are meant to pick out. For others, resisting marginal concepts and conceptions additionally involves the construction of mistaken conceptions, resulting in conceptual non-competence. For example, one prominent misconception concerning trans people is that they are ‘confused or actively misleading others by identifying with a gender different from the one assigned to them at birth’ (Lopez 2017). In either case, these agents are responsible for their conceptual ignorance or non-competence. The reasons-responsive cause of their resistance to these concepts and conceptions — and of their ignorance that they are wrongly remaining conceptually ignorant or non-competent — is revealed by the very rationales they present for resisting marginal concepts and conceptions.

The agent need not actively be resisting marginal concepts and conceptions to be culpable for her conceptual ignorance or non-competence, however. Implicit prejudice against hermeneutically marginalized social groups may influence the agent’s perception of speakers from those social groups without her conscious awareness, leading her to view marginalized speakers as less credible than they are. With regard to testimony of propositional knowledge, Fricker identifies such prejudicial credibility

deficits as instances of *testimonial injustice*, another kind of epistemic injustice (Fricker 2007). Testimonial injustice may also undermine hermeneutically marginalized speakers' credibility with regard to the concepts and conceptions they have developed for their distinctive experiences.<sup>89</sup> The hearer's already existing (though not explicit) prejudice against the speaker compounds the communicative unintelligibility of the hermeneutically dissenting concepts used by the speaker, making it all the more unlikely that the hearer will consider the possibility that the speaker is not spouting nonsense. As Medina puts it, 'testimonial insensitivities and hermeneutical insensitivities feed each other' (Medina 2013, 96). Where this kind of testimonial injustice serves to maintain the hearer's conceptual ignorance or non-competence, she is responsible for her conceptual ignorance or non-competence, because her conceptual incapacities and her ignorance of wrongdoing are caused by a reasons-responsive but misdirected psychological mechanism of hers.

However, there are cases where we may hesitate to attribute the agent's conceptual ignorance or non-competence to her, on the basis of widespread cultural ignorance. In cases where hermeneutical marginalization is severe enough, very few people outside of hermeneutically dissenting communities will have mastered the concepts and conceptions needed to understand certain marginalized subjects' experiences. We might construe this situation as one where evidence is simply not available to the relatively dominant subjects who are conceptually ignorant. They would then fail condition (E2), and be excused for their conceptual ignorance. It may be, for example, that until recently the marginalization of LGBTQ+ communities prevented many people from even hearing about some of their distinctive experiences, much less the concepts needed to understand them.

---

<sup>89</sup> Cf. Derek Anderson's account of *conceptual competence injustice*, an epistemic injustice where the agent is unfairly regarded as failing to grasp the concepts she employs (D. E. Anderson 2017)

But, recall the point about cultural ignorance I raised at the end of the last chapter: Elinor Mason and Alan Wilson argue that cultural ignorance does not provide an excuse as often as many have claimed, because in many instances the evidence *was* available, but was not properly received because of widespread intellectual *vices*, especially epistemic arrogance and epistemic laziness (E. Mason and Wilson 2017). Arguably, evidence has been plainly available in many instances of cultural ignorance — of the wrongfulness of slavery in the 18<sup>th</sup> century, the wrongfulness of sexism in the 20<sup>th</sup> century, perhaps the wrongfulness of meat-eating now — but most people have been too lazy or too arrogant or simply too indifferent to consider inquiring.<sup>90</sup> We might say the same thing about certain concepts and conceptions. Perhaps there was ample evidence available to sexual harassers in the 1950s that their behaviour was not just improper but wrongful, and their conceptual ignorance of something like the concept SEXUAL HARASSMENT was due not to a lack of evidence but to their intellectual vices. And, on the view of responsibility I have defended, ignorance that is the result of vice is not an excuse, because it is caused by a reasons-responsive but misdirected psychological mechanism.

If Medina (2013) is correct and many of these vices are common among privileged social groups, then a great many cases of conceptual ignorance will be culpable, despite being widespread. Still, there genuinely are cases where the agent is excused for her conceptual ignorance by a lack of evidence. Mason and Wilson's argument does not refute the possibility of cultural ignorance; it only serves to undermine many purported cases. Sometimes, the agent may be conceptually ignorant because of bad circumstantial luck. An ancient Greek, for example, might seem unlikely to have access to the information required to arrive at a concept of mental illness. We might think that the cultural environment, the range of experiences had in that setting,

---

<sup>90</sup> On indifference to the truth as an epistemic vice, see Quassim Cassam's account of 'epistemic insouciance' (Cassam, forthcoming).

the background assumptions about the mind and religion, and so on, simply make it impossible, or highly unlikely, for that conceptual development to take place. Where environmental factors genuinely cut the agent off from information needed to acquire certain concepts, she is not responsible for her conceptual ignorance.

### 3.8. Summary

In this chapter, I have argued that we are sometimes responsible for our conceptual incapacities, in the sense that they can be attributed to us such that we are open to moral appraisal for using problematic concepts or being conceptually ignorant or non-competent. After examining multiple theories of responsibility, I developed a view I called reasonable psychologism, and demonstrated that, on this view, we are responsible for our conceptual incapacities in a wide range of cases. Even if reasonable psychologism is wrong, however, we can still hold that we are sometimes responsible for our conceptual incapacities. For, I argued that each of the other views I canvassed in §§3.2–5 enable us to judge that, in at least some cases, we can attribute the agent's conceptual incapacities to her. In Chapter 4, I argue that, regardless of whether the agent turns out to be responsible for her conceptual incapacities, she is still *answerable* for them, in the expanded sense I outlined in §3.1, and thus, morally and epistemically required to *take* responsibility for them.

#### 4. Conceptual Answerability

In the first two chapters, I argued that our conceptual incapacities — our use of, ignorance of, or non-competence with certain concepts — can be morally, epistemically, or socially problematic. In the previous chapter, I argued that we can, at least sometimes, be morally responsible for our conceptual incapacities. Still, in a significant range of cases, the agent will not be responsible for her conceptual incapacities. These may include cases where the appropriate concepts and conceptions are not in use by anyone, or are possessed only by the severely hermeneutically marginalized. Such cases plausibly represent instances where evidence is simply not available to the agent that her concepts are problematic, thereby excusing her from attributions of responsibility for her conceptual incapacities. Regardless, whether or not we are responsible for our conceptual incapacities, it remains unclear just what we are required to do in light of their pernicious effects. In this chapter, I argue that when the agent is at fault for something — and in an important range of cases where she is not at fault for something but still closely connected to it — the agent is morally required to take steps to *answer* for that thing. In the case of our conceptual incapacities, I argue that the required response is to critically reflect upon our concepts; the next chapter expands upon the nature of this reflection.

In the first section, I distinguish between the moral notions of taking responsibility for something and being responsible for something. In the central cases, the former concerns what one is morally required to do to answer for some fault. But, as I argue in the second section, there are a range of cases where the agent may be required to take responsibility for something without being at fault for it. These scenarios include bad moral luck, and events that fall under what David Enoch calls our *penumbral agency*. In the third section, I return to conceptual incapacities, arguing that both in the cases discussed in the previous chapter (where the agent is responsible for her conceptual incapacities), and in a wider range of cases (where the agent is morally

and epistemically innocent for her conceptual incapacities), the agent is morally and epistemically required to take responsibility for her concepts. The central response that is required of her is to engage in critical reflection; I introduce a sketch of this activity as presented by Bernard Williams, to be filled in with more detail in the next chapter.

#### 4.1. Taking Responsibility and Being Responsible

In the previous chapter, I distinguished between the notions of being responsible for something, and taking responsibility for something. For the agent to be responsible for something, that thing must be properly attributable to her such that she is open to moral appraisal for it. The bulk of Chapter 3 consisted in defending my preferred account of attributability, reasonable psychologism, according to which an action or mental reaction is properly attributable to the agent just in case it was caused by one of the agent's constitutive reasons-responsive psychological traits, and the agent was in a position to know that it was wrong. In the course of that discussion, I briefly noted that the reasons-responsiveness of those traits is what opens the agent to demands to provide her justificatory reasons, if any, for her behaviour. This demand is a request that the agent make explicit her values or evaluative framework, explaining her reasons for taking the action, acquiring the attitude, or reacting as she did. Where the agent has acted wrongly, answering this demand may reveal her to be irrational, morally ignorant, morally justified, or else will reveal an excuse. Being open to this kind of justificatory demand is emphasized by philosophers such as Angela Smith (2005, 2012), Marina Oshana (1997), and Pamela Hieronymi (2014) as the condition for being *answerable* for one's behaviour.

But as I outlined in the previous chapter, the notion of being open to demands to answer for something is broader than being open to demands for justifications. Answering for something primarily concerns what the agent is morally required to do

by virtue of being responsible for something.<sup>91</sup> And the range of responses we expect of agents who are at fault for something is much wider than just attempting to justify themselves when asked. For instance, we typically take it that the agent owes those affected by her actions an apology. If I carelessly remark that a student's work is unusually poor and hurt the student's feelings, I should apologize. Where the agent's wrongdoing is rooted in a character flaw, the agent may be required to undertake a project of self-improvement to prevent similar acts in the future. If you and I are both sitting on a disciplinary committee and you ask me to back you up on pushing for sanctions against a department member, but I chicken out for fear of reprisal from the administration, I owe it to you not just to apologize for my cowardice but also to work at being more courageous in the future. Where one's actions have done material harm to one or more victims, one may be morally required to compensate those affected. If I recklessly kick a football through your window, I owe it to you to cover the replacement of the glass, either by paying you the cost, doing it myself for free, or by offering some equivalent service, say, weeding your garden or shovelling the snow off your sidewalk. Depending on the severity of the moral infraction for which one is at fault, the response required may be relatively easy to fulfil (a gift of flowers may suffice for forgetting a partner's birthday), or it may require more than one can do in a lifetime (many years of community service may not shed the burden imposed by having committed war crimes). I suggest that this wider sense of answering for one's actions aligns with the primary notion of *taking responsibility* for something. That is to say, to answer for something is to take responsibility for it, which is to fulfil moral obligations that one has incurred regarding that thing by virtue of one's connection to it —

---

<sup>91</sup> In the following, I will repeatedly refer to the obligations and moral requirements of being answerable for something — our 'answerability-obligations'. The notion of moral obligations is, of course, controversial (cf. Anscombe 1958a). But nothing in particular hangs on the notion of *obligation* in my account. If you prefer, you may think of being answerable for something not as generating obligations to respond in a particular kind of way, but as giving one *reasons* to respond in a particular kind of way.

paradigmatically, by being at fault for it. To be open to demands of this sort is, I contend, to be answerable for it.

It may be argued that while we may be required to take certain actions in response to our wrongdoing, this is not the notion captured by the phrase 'taking responsibility'. Rather, the objection goes, what we mean when we say that someone is taking responsibility for something for which she is at fault, we are saying that she is *admitting* that fault. To take responsibility, on this interpretation, is to 'own up' to what one has done. David Enoch, for instance, claims that 'If I say, "It's time that you start taking responsibility for your actions!" ... What I typically have in mind is that it's time that you acknowledge or recognize the relevant responsibility-facts, namely, that you are responsible for your actions' (Enoch 2012, 103–4). On this view, then, the demand for the agent to take responsibility for her actions is a *moral-epistemic* one, not a *practical* one.

But while this moral-epistemic aspect is important, Enoch is mistaken to identify it as the central activity of taking responsibility for some fault. Consider the following example. Jackie, a black woman, is out drinking with co-workers, including Hank, a white man. After having a little too much to drink, Hank tells a racist joke. Suppose that Hank's inebriation neither excuses nor aggravates his having told the joke, and that he is not excused by virtue of non-culpable ignorance or any other factor. In other words, whether he realizes or not, Hank is morally responsible for insulting Jackie. When next she sees him at the office, Jackie may confront Hank with the phrase, 'You need to take responsibility for your actions!' or, perhaps more directly, 'You need to apologize for what you said.' In saying this, Jackie is demanding the apology Hank owes her in virtue of being responsible for insulting her the previous night. Jackie is not just trying to make Hank *aware* that he did something wrong or to get him to *admit* to or *acknowledge* his wrongdoing. While it is important that he own up to his wrongdoing as part of taking responsibility for it, this in itself is insufficient. For he



could admit responsibility, and even own up to wrongdoing, without then apologizing or doing anything else that he owes her. (Consider the possibility that he admits to having done something wrong and gloats about it, taking pride in being ‘transgressive’ and ‘politically incorrect’.) That Jackie might have more directly confronted Hank by demanding an apology outright suggests that injunctions to ‘take responsibility’ for something involve more than just an epistemic demand. Rather, demands that the agent take responsibility also aim to spur him to do what he is morally required to do in virtue of being responsible, that is, *to fulfil his answerability-obligations*. An admission of wrongdoing only partially fulfils these obligations. Making this demand may also serve the moral-epistemic function of making the agent aware of his responsibility for a wrong he has committed, but only because the agent may sometimes be responsible regardless of his own awareness of having done wrong, as in cases of culpable ignorance.

This notion of taking responsibility for one’s actions as taking the actions one is required to take in virtue of being answerable for those initial actions also receives support from situations in which we refer to ourselves as ‘taking responsibility’ for something. If I write an apology to a friend I have wronged, one thing I may say to explain why I have written is that I have decided it’s time I took responsibility — or answered — for what I did. Again, the point is not primarily moral-epistemic: in saying that I am taking responsibility for wronging my friend, I am indicating not merely that I acknowledge my wrongdoing. Rather, I am signalling the sincerity of my apology and my willingness to make amends. Similarly, suppose you pull a prank on a roommate, using the old bucket above the door trap. The trap springs successfully, but in the process the water spilling from the bucket wrecks your roommate’s laptop computer, which he was carrying at the time. If you make amends for your bad-joke-gone-wrong by buying your roommate a new computer, you might describe this compensation as taking responsibility for your actions. Again, the point is not primarily epistemic: you

are not simply acknowledging your responsibility for the prank, you are in the process of fulfilling a moral requirement to compensate your roommate for the loss caused by your recklessness.

The notion of taking responsibility for something, then, is best analysed as *taking the actions one is morally required to take in virtue of being answerable for something*. Where one is at fault for that thing, the required actions often include admitting that fault, but they typically include a variety of other actions that one owes to those affected by one's actions: apologizing, making amends, improving one's character, and so on. Since, as I argued in Chapter 3, we may sometimes be at fault for our conceptual incapacities, the question is then: what are we required to do to take responsibility for them? But before tackling this question, it is worth returning to a point raised at the end of the last chapter, namely, the issue of bad moral or epistemic luck. As I noted, it is plausible that ignorance sometimes excuses us from fault for our conceptual incapacities. Would that mean that, in these cases, we are not required to take responsibility? On the contrary, I think it is plausible that, in cases of bad luck and some similar scenarios, the agent may be morally required to take responsibility for things that are *not* properly attributable to her. In other words, the agent may be answerable for certain things without being at fault for them. If this is so, then even in cases where the agent is excused from fault for her conceptual incapacities due to cultural ignorance, mental illness, or some other factor, she may still be required to take responsibility for her concepts.<sup>92</sup>

---

<sup>92</sup> Moreover, if it can be shown that we may be answerable without fault, this establishes a case that we are required to take responsibility for our concepts even if the account of moral responsibility I defended in the previous chapter is mistaken. If, say, voluntarism is correct and we are hardly ever responsible for our conceptual incapacities, on the view defended in this chapter we would still be answerable for them, and by that token required to take responsibility for them.

#### 4.2. Bad Luck and Penumbral Agency

To understand the import of bad luck for conceptual responsibility, it will be instructive to consider the issue of bad luck with regard to moral responsibility more generally. Bad moral luck occurs when circumstances outside one's agency seem to make the one morally responsible for some event despite not being at fault for it.<sup>93</sup> This gives rise to two competing intuitions. On the one hand, it seems wrong to judge that the agent is responsible for an event for which she is not at fault, even if it is connected to her somehow. On the other hand, there are cases where we expect and even demand that the agent respond in a morally significant way to such events, in a manner different from that expected of a bystander. For example, Bernard Williams describes the case of a truck driver who, through no fault of his own, has run over a child (Williams 1981). Following the first intuition, we are led to the conclusion that the driver is in no way responsible for the accident, so nothing can be demanded of him in virtue of having caused the accident. But this is in tension with the second intuition: Williams observes that we in fact *do* feel there is something the driver should do. Namely, we expect him to feel 'agent-regret': he should feel bad that *his* actions led to a tragic result, wish that he had acted differently, and desire to make amends. This kind of regret is different from that which may be felt by a third party observer. Both the agent and the observer may feel sad that things turned out differently and wish that things had gone otherwise, but only the agent can have that sense that *if only he had acted differently*, the tragedy could have been avoided. As Williams remarks, 'there is something special about [the driver's] relation to [the accident], something which

---

<sup>93</sup> Strictly speaking, moral luck is a broader phenomenon than this, encompassing any instances where luck seems to make the agent an appropriate target of moral judgement (Nelkin 2013). The kinds of moral luck discussed by Nagel (1979), for example, include 'constitutive luck', where the agent's moral character is largely a result of her genetics and how she was raised, rather than anything she is plausibly responsible for, yet still seems to be a proper target of moral appraisal. I have limited my discussion of moral luck to the variations that are most important with regard to our conceptual incapacities.

cannot merely be eliminated by the consideration that it was not his fault' (28). It is this kind of connection to the unlucky outcome that generates the sense that the driver, and others in similar positions, ought to feel and act in a way distinctly different both from the responses of an uninvolved observer and an agent who intentionally caused the bad outcome.

Enoch denies that bad moral luck can make you responsible, in any sense, for things that cannot be attributed to your agency.<sup>94</sup> Rather, he suggests that 'you are not responsible... you are rather under a moral duty to take responsibility' (Enoch 2012, 100).<sup>95</sup> Enoch's sense of 'taking responsibility', here, is not the same as mine, and is different from the sense of acknowledging fault mentioned above. The relevant notion is rather what we may call *assuming responsibility*.<sup>96</sup> When you assume responsibility for something, you perform an act of will to change the responsibility-facts. Assuming responsibility *cannot* change the responsibility-facts that concern proper *attributability*; nothing you can do after the fact will change whether an action or outcome can be attributed to you. Rather, assuming responsibility for something means taking on the status of being *answerable* for it, in my sense, acquiring in the process all the associated obligations to, as I put it, *take* responsibility for it: to justify, to apologize, to make amends, and so on.

To illustrate Enoch's proposal, consider the truck driver. He is not at fault for the accident — he was driving perfectly competently and carefully, and the child just happened to dash out into the street too late for him to brake or swerve. However, as

---

<sup>94</sup> See Enoch and Marmor (2007).

<sup>95</sup> Enoch does not endorse this claim outright. He rather says that he intends to 'defend the claim as a coherent conjunction' (Enoch 2012, 100); that is to say, his aim is to establish room in logical space for this kind of position. His aim is limited because too many of the other details of his account concerning the nature of responsibility and the aforementioned 'special relation' between agent and penumbral agency (see below) remain sketchy.

<sup>96</sup> In the quotations from Enoch (2012) below, I have accordingly replaced instances of 'take responsibility' and similar with 'assume responsibility' and similar.

Williams observes, because of the role he (faultlessly) played in causing the accident, we expect the driver to feel agent-regret and to attempt to set things right. On Enoch's view, this expectation is cashed out as an obligation on the driver's part to assume responsibility for the tragic event. Once the driver assumes responsibility for the accident, he incurs further obligations to feel regret and to visit the injured child multiple times in the hospital to apologize. Enoch concludes that his proposal provides a plausible way of 'accommodating the agent-regret intuition apparently without conceding the existence of moral luck' (Enoch 2012, 130).

However, although I acknowledge that there are cases where we may assume responsibility for events outside the scope of what may be attributable to us, I think it is a mistake to hold that this step is required before being answerable for the outcomes of bad moral luck.<sup>97</sup> That is to say, I think it is more plausible that the agent is already morally required to take responsibility, as I defined it, for the outcomes of bad moral luck, without first having to assume responsibility for those outcomes. In the truck driver example, notice that we expect the driver to feel agent-regret and to want to

---

<sup>97</sup> Assuming responsibility for something, I think, concerns rather a range of events that are connected to our agency only tenuously, if at all. For instance, we assume responsibility for something whenever we take on some role or duty (and this indeed coincides with another sense of the phrase 'taking responsibility'). While it is plausible that we may have obligations to assume responsibility for such things, they are not of the sort I am concerned with here. In fact, I suspect that in many instances it will be either inappropriate or supererogatory to assume responsibility for various things. Film and TV dramas often feature such examples; in most cases another character who is more closely involved in the matter will confront the character who assumes responsibility by saying something like, 'I never wanted you to get involved!' or 'It's none of your business, so stay out of it!' In some cases, the character taking responsibility is portrayed as doing something good that goes beyond her obligations. But in other cases, the assumption of responsibility was an invasive overreach. Consider the following case from the police procedural 'Rosewood' (Harthan 2015). The protagonist, Dr. Beaumont 'Rosie' Rosewood, a private pathologist hired by the Miami Police Department to assist in forensic investigations, secretly investigates the death of the husband of Detective Annalise Villa, a colleague with whom he is infatuated. When he reveals his investigation to Villa, telling her that he thinks her husband may have been murdered, she reproaches him for violating her privacy. Rosie's decision to open an investigation without Villa's knowledge or permission is plausibly a case of inappropriately assuming responsibility where none existed for him.

make amends straightaway. It would strike us as very odd and slightly disturbing if the driver were coolly indifferent to the accident *for any length of time* on the basis that he was not at fault. Enoch claims that our disapprobation of the indifferent driver is grounded in his failure to fulfil the obligation to assume responsibility for the accident, for were he to assume responsibility for the accident, he would *then* be morally required to feel moral agent-regret. But this leads to an odd consequence: unless and until the driver in fact assumes responsibility for the accident, we cannot criticize him for not feeling agent-regret or for not apologizing. This is because, on Enoch's view, the driver is not yet responsible in any sense for the accident, so *a fortiori* not answerable for the accident. Therefore, the driver who has not yet assumed responsibility for the accident is under no obligation at all to feel regret or to apologize or otherwise make amends for the accident. It may be appropriate to reproach the driver for not assuming responsibility for the accident, but it would be unfair to reproach him for not doing the things he would be (*but is not now*) obligated to do were he to assume responsibility.

Compare a case Enoch uses to set up the plausibility of the obligation to assume responsibility for the outcomes of bad moral luck. Enoch argues that there are instances where we are morally required to make certain *promises*. He asks us to imagine that a friend of his, who is near death, is concerned that his young daughter will lack a caregiver after he dies. They are not sufficiently close friends for Enoch already to have an obligation to take care of his friend's daughter after his death, and no one else has promised to care for her. Enoch suggests that 'there are ways of filling in the details of the example such that I shall be morally required to promise my friend to take care of his daughter if he dies' (Enoch 2012, 105). Nevertheless, should he fail to make this promise, he is under no obligation to care for his friend's daughter after the friend's death — he would not, therefore, be at fault for failing to care for his friend's daughter, only for failing to *promise* to care for his friend's daughter. And so, it would not be appropriate to criticize Enoch for not caring for his friend's daughter, only for

failing to promise to do so. On Enoch's view, it would be just as wrongheaded to fault the indifferent driver for failing to feel agent-regret and for failing to apologize to the child. Because the driver never assumed responsibility for the accident, he was *never* morally required to feel regret or to apologize for the accident, and he should not be criticized for failing to meet obligations he does not have.

I submit that Enoch's account of the driver's obligations regarding the accident is needlessly circuitous and out of step with our actual practices. If, as Enoch proposes, the indifferent driver must first assume responsibility before having the obligation to feel agent-regret or to apologize to the child, then we must hold back in our demands and reprimands; we can find fault for his failure to take the mental act of assuming responsibility and no more. But on the contrary, I think it is more plausible that the driver *already has obligations* to feel agent-regret and to visit the child to apologize. Just as we would demand that he take responsibility for his actions if he were at fault for the accident, we may demand that he take responsibility for the accident in the case of bad moral luck — though precisely what we may demand of him in each case may be different. In other words, *despite not being at fault, the driver is answerable for the accident*. In the case of the indifferent driver who fails to feel regret and to apologize for the accident, the disapprobation we feel towards him is directed precisely at his failure to feel an appropriate kind of regret and to apologize, not for failing to assume the obligation to feel such regret and to apologize. He already has this obligation, because he is already answerable.

Enoch anticipates my proposal, and tries to head it off. He remarks that on his account, 'one can implicitly [assume] responsibility by performing the first part of what it would take to live up to the responsibility just [assumed]' (Enoch 2012, 110). That is to say, on his view, the mental act of assuming responsibility may coincide with taking one of the actions that one would be obligated to perform by virtue of one's having assumed responsibility. For example, should the driver immediately feel regret and

take steps to apologize and make amends, on Enoch's view this would be to assume responsibility and to fulfil the obligation to take responsibility thus incurred at one stroke. But, as Enoch admits, this makes the logically prior step of assuming responsibility look redundant. So why not just say, as I am suggesting, that we are already answerable for the outcomes of bad moral luck?

Enoch claims that there is an important moral difference between someone who, on his account, *fails to assume responsibility*, and someone who *fails to live up to the responsibility assumed*. He uses the truck driver example to make this point. Suppose that part of what is morally required of the driver, should he be answerable for the accident, is to *repeatedly* visit the injured child in the hospital to apologize. It seems there is a distinction between the moral fault of Driver A, who *never* visits the child, and Driver B, who visits *once* but never again. On Enoch's account, there is a clear difference in the kind of fault between the two cases:

Not visiting at all amounts... to a violation of the moral duty to [assume] responsibility. Coming once may well amount (depending on the circumstances) to [assuming] responsibility. From that point on, not coming again amounts to failing to respond appropriately to the fact that one is (now) responsible for the injury. (Enoch 2012, 111)

Enoch claims that the sort of theory I am proposing — where the driver was answerable for the injury all along — sees no difference between the cases. His argument seems to be that on my view, both drivers fail to live up to the *same* obligations, while on his view the two drivers fail to live up to *different* sets of obligations.

But on the view that I propose, we can still make a distinction between the two drivers. Let us suppose with Enoch that the drivers are morally required to visit the injured child multiple times in the hospital. On my view, they already have this obligation in virtue of being answerable for the accident; no intermediate step of assuming responsibility is necessary. One way we could make the distinction between the two drivers is in terms of their moral beliefs. Driver A, who never visits the child,



seems to fail to realize (or to actively deny) that he is answerable for the accident. Driver B, who visits once and never again, demonstrates that he knows he is answerable for the accident, yet he fails to act appropriately (i.e. to visit multiple times), perhaps because he is akratic and fails to visit again for fear of shame, or perhaps he fails to recognize the full extent of his answerability-obligations. Another way to make the distinction is in terms of the extent to which the two drivers fail to fulfil their obligations. Driver A, in not even visiting once, shirks *all* of his obligations in virtue of being answerable for the accident. Driver B, in visiting the first time, *partially* fulfils his obligations, but, by not returning again, flagrantly shirks the rest. Another way to put it is with regard to how each driver fails properly to take responsibility, as I defined it, for the accident. Driver A *fails to take responsibility* for the accident, that is to say, he fails to act in accordance with a moral requirements he has in virtue of being answerable for the accident. Driver B *starts to take responsibility but does not follow through*, that is to say, he only partially fulfils the moral requirements he has in virtue of being answerable for the accident. Whichever way we put it, we can still make a morally important distinction between the two drivers. So, Enoch has not demonstrated the preferability of his proposal.

As mentioned, Enoch's proposal is motivated by a denial that bad moral luck can make us responsible, in any sense, for the tragic outcomes of our actions. It may then be suggested by way of reply to my own proposal that I have just sided with the opposite intuition, namely, that agent-regret and similar phenomena show that bad moral luck *does* make us responsible in these cases. But far from simply taking one horn of the dilemma, I contend that my account enables us to *dissolve* the conflict between the two intuitions — not so much slipping between the horns of the beast as revealing the creature to have been a phantasm all along. For, if I am right, cases of bad moral luck are instances where we are answerable for happenings connected to our agency in a way weaker than attributability. If we decouple attributability and (my

sense of ) answerability, we can affirm the existence of moral luck without admitting that we are responsible for its effects in any problematic sense — that is, without admitting that we are *at fault* for these outcomes. Without making space for answerability without attributability, moral luck can seem puzzling, but once we acknowledge that the former is not a sufficient condition for the latter, the dilemma vanishes.

Still, this raises the question of how we may be answerable for events that are not attributable to our agency. To see how this works, it is useful to consider Enoch's wider purpose in discussing moral luck and assumptions of responsibility. On his view, there is a broad range of phenomena where a special relation other than fault obtains between our agency and morally problematic events. In addition to Williams's truck driver case, Enoch provides two further examples. The first concerns actions taken by close others: 'Your teenage son commits a crime, causing harm to person and property. You are not, let us suppose, directly responsible for the crime in any straightforward way — it's not as if you put him up to it, or even drove him to this kind of thing by your poor parenting' (Enoch 2012, 97). Despite not being at fault for your son's actions, however, Enoch notes that there seems to be something morally wrong with *saying* so, particularly where this is an excuse for not taking some further action, e.g. apologizing for what your son did (though perhaps not making an apology *on his behalf*).<sup>98</sup> Enoch's second example concerns actions of a collective of which one is a member, for instance, when a citizen's (legitimately governed and democratic) country wages an unjust war. Even supposing the citizen was not in any way at fault for his government's decision, when his country's actions face moral criticism 'there would be something wrong... if all he did was to (correctly) point out that these actions are not his... it's not that this

---

<sup>98</sup> If the intuition seems weak here, suppose that the son is a rather young teenager, say, fourteen years old. There is thus also a legal case for taking responsibility for his criminal behaviour.

would be false... Rather, it's that his response should at the very least be more complicated than this', involving, perhaps, a justification of, apology for, or denouncement of his country's actions (98).<sup>99</sup> Enoch groups these cases, along with bad moral luck, under the label *penumbral agency*. The image is that our presence in the world casts a shadow, with events for which we are at fault lying in the darkest part (our 'core' or, to extend Enoch's image, 'umbral' agency) and those to which we have some special moral relation other than fault lying in the lighter shades at the edges (our 'peripheral' or 'penumbral' agency).

What unifies these various cases under the category of penumbral agency is roughly that these events each invoke, as Enoch puts it, 'ways in which we are morally required to think of ourselves in a normatively rich kind of way' (Enoch 2012, 128). That is to say, there are certain self-descriptions, self-conceptions, personal commitments, or other aspects of our personal identities that are, in certain circumstances, particularly morally relevant. Being a parent and being a citizen are morally important aspects of our identities that generate obligations, among which, for Enoch, are obligations to assume responsibility for the actions of one's children and one's country, and thereby to acquire further answerability-obligations. With regard to the harmful outcomes of our actions that are due to bad moral luck, the relevant aspect of our identities is more subtle. As Enoch puts it, 'perhaps we are also morally required to think of ourselves as *limited*, as creatures who have some but far from full control over the consequences of their actions in the world' (127).<sup>100</sup> Conceiving of ourselves in

---

<sup>99</sup> If the intuition in this case seems weak, suppose that the citizen disapproves of the war, but is usually proud of his country's achievements and even volunteered for the governing party in the previous election campaign. His country's actions are still not his own, but he seems more strongly implicated if we fill in these details.

<sup>100</sup> In a similar vein, Susan Wolf writes that taking responsibility for the outcomes of bad moral luck expresses 'our recognition that we are beings who are thoroughly in-the-world, in interaction with others whose movements and thoughts we cannot fully control, and whom we affect and are affected by accidentally as well as intentionally, involuntarily, unwittingly, inescapably, as well as voluntarily and deliberately' (Wolf 2001, 14).

this light makes the obligation to assume responsibility for some of the regrettable but unforeseen or unavoidable consequences of our actions seem plausible, especially when we notice that we often take credit for *good* outcomes of our actions that are a matter of luck.<sup>101</sup>

Of course, I am going to say there is no need to assume responsibility, as we are already answerable for events in the penumbra of our agency. But, I think what accounts for this answerability without fault are precisely the same morally significant aspects of our identities that Enoch appeals to in grounding the obligation to assume responsibility. These aspects of our identities generate reasons for acting in certain ways — in cases of penumbral agency, reasons to take responsibility — which, as moral agents, we should be sensitive to. The fact that we should be responsive to these reasons is what generates both our answerability-obligations and the appropriateness of others making demands that we take responsibility for these events. Similarly, in cases of fault, our relationship to the behaviour which can be properly attributed to us generates reasons to take responsibility for our actions. In this case, the way in which we are required to think of ourselves is as *moral agents*. That our reasons-responsive constitutive traits — those of our traits that are both part of what makes us who we are and part of what account for our ability to act in the world as agents — have failed to operate as they should reflects badly on us. And for this, we may appropriately be asked to make amends.

---

<sup>101</sup> Exactly how far beyond the immediate consequences of our actions the shadow of answerability extends is difficult to make out, and the boundaries are likely to be vague. But the issue is not fundamentally different from a more general problem of determining the extent of the consequences of our actions for which we are responsible in the core sense. One provisional answer that seems plausible enough is that the agent is answerable for those of the unlucky consequences of her action for which she would have been responsible were she at fault for her action. Whatever principle limits us from being responsible for all of the inordinate consequences that radiate out from those of our actions for which we are responsible also limits the consequences of our actions for which we are answerable but not at fault.

One might object that we can just as well account for our sense that the truck driver and the others should ‘answer’ for events in the penumbra of their agency by noting that they are *responsible for how they respond to these events*. It is their responsibility at this later stage, and not their answerability for the earlier events, that accounts for our disapprobation where they fail to respond in the appropriate way. We are, after all, routinely morally required to respond to various events that have nothing to do with our agency — core or penumbral. If I happen across a drowning man, for example, I am obligated to try and save him, and if I do not I am liable to be blamed, but it does not seem right to say that I am *answerable* for his predicament or required to *take responsibility* for it. We might say that penumbral agency is similar: the agent is morally required to respond to these events, but is still not answerable for them. It is only a coincidence that the required response resembles answerability-obligations.

But I think this objection misses the point. It is an open question whether the parent, the citizen, the driver, or agents in similar situations are responsible for how they respond. Suppose the parent suffers amnesia shortly after learning about her son’s crime, for example, and forgets the last three months. We would not then judge that she is responsible for failing to take responsibility for her son’s actions, because she is no longer in a position to be aware that he committed a crime. But it still seems true that by virtue of being a parent she is in a special relation to her son’s actions, such that she would be required to take certain actions, such as apologizing, in response, should she re-learn of her son’s actions. Moreover, these obligations are of distinctive genus that attaches to answerability — apologizing, acknowledging fault if appropriate, seeking to make amends, and so on — and not just instances of moral obligation more generally. After saving the drowning man, as an intervening bystander I would not be under any moral requirement to apologize to him, though the person who owns the pool into which he fell might be.

To sum up my proposal, I have argued that we are already answerable for events in the penumbra of our agency. These events include the outcomes of our actions which are due to bad moral luck, and for which we are, thus, not responsible in the core sense. That we may be answerable for these events without their being attributable to us allows us to dissolve the problem of moral luck by acknowledging the insights of both of the competing intuitions. Enoch and others who deny the existence of moral luck are right that these outcomes cannot be properly attributed to our agency. But Williams and others who accept the existence of moral luck are also right that we are responsible in some sense for these outcomes.<sup>102</sup> Once we see that we may be answerable for things that are not attributable to us, we can accept both sides of the debate.

#### 4.3. Moral and Epistemic Answerability

A great deal of our conceptual repertoire is the result of luck. As discussed in Chapter 1, we acquire most of our concepts, notably our evaluative concepts, through being brought up into a set of cultural practices that form the background against which those concepts make distinctions and evaluations. As John McDowell remarks, ‘a natural language... serves as a repository of historically accumulated wisdom about what is a reason for what’ (McDowell 1996, 126). But as we have seen, not all is well with the concepts and conceptions that we inherit in this way. As Matthew Congdon observes, ‘It is not only the wisdom of past generations that is memorialized in language, but also the folly, the blind spots, the prejudice, the influence of abused power’ (Congdon 2015, 86), which can result in our having conceptual incapacities. The kind of cultural environment in which we are raised is largely a matter of luck: we are

---

<sup>102</sup> Whether Williams actually affirms that bad moral luck can make us responsible is disputed. Some have pointed to a postscript to his initial essay as evidence that he did not: there he writes that he intended the phrase ‘moral luck’ to suggest an ‘oxymoron’, in service of his more general project of extending our conception of ethics beyond what he later called ‘the moral system’ (Williams 1993).

not responsible for who our parents are, where we are born, or the surrounding society. To some extent, this lack of responsibility for our cultural environment continues into adult life: the sorts of factors that might expose us to new information that introduce us to new concepts and conceptions or that challenge our received conceptual repertoire — which narratives and voices become popular, the opportunities we have to change our social conditions, and which hermeneutically marginalized perspectives we happen to encounter — are all significantly a matter of luck, despite our increased ability, as adults, to intervene in our own cognitive development. We should therefore expect that there will be a wide range of cases in which our conceptual incapacities are due more to our unlucky cultural ignorance rather than to any fault on our part.

Building on the case made in the previous section that we may be morally answerable (in my sense) for events that are not attributable to our agency, I want to argue that much, perhaps all, of our cognition lies within either the core or the penumbra of our agency. Towards the end of Chapter 3, I outlined several cases where we may be at fault for our conceptual incapacities: the deliberate use of a problematic concept, the unconscious or unreflective use of a problematic concept, actively remaining conceptually ignorant or non-competent in opposition to hermeneutically dissenting communities, unconsciously remaining conceptually ignorant or non-competent due to implicit prejudice, and remaining conceptually ignorant or non-competent due to intellectual vice. These cases are examples where our conceptual practices fall squarely in the core of our agency, so in these instances we are morally answerable for our conceptual incapacities. In other cases, such as non-vicious cultural ignorance, our conceptual incapacities are not attributable to our core agency. But on the account of penumbral agency I propose above, according to which may be answerable for events not attributable to us, we are still morally answerable for our

conceptual incapacities. To see why, it is useful to compare the moral notion of answerability with an epistemic analogue.

Few would contest the idea that we are, in general, *epistemically* answerable for our *beliefs*. By ‘epistemically answerable’ I mean something analogous to answerability in the moral case: the subject is epistemically required to take responsibility for her beliefs by, for instance, justifying what she believes when requested. To take a toy example, if I believe that my mug is on my desk, I am open, in principle, to a request to provide my justificatory reasons for believing that my mug is on my desk, namely, that I can see it there. Or to take a more complex example, when Donald Trump says that ‘The concept of global warming was created by and for the Chinese in order to make U.S. manufacturing non-competitive’ (@realdonaldtrump 2012), he is open to demands that he provide his justification for believing this conspiracy theory.<sup>103</sup> Moreover, just as moral answerability requires that the agent take responsibility by taking actions responding to that for which she is answerable (such as offering an apology), epistemic answerability may require the agent to change her beliefs and possibly to take further actions to remedy any (morally or epistemically) harmful effects of her error. If it turns out that my mug is not on my desk — what I thought was my mug is really a different token of the same mass-produced type — then it turns out that my justification was spurious, and I am epistemically required to change my belief. When it turns out that Trump cannot provide any sound justification for his climate change denial, he is epistemically required to take responsibility for his error by changing his beliefs — additionally, it seems right that given his influence and his social position, he should also be required (epistemically and morally) to publicly admit the error and issue a correction.

---

<sup>103</sup> Maybe Trump does not actually believe half of the strange and dangerous things he says, and simply does not care about the truth when he speaks, like Frankfurt’s *bullshitter* (Frankfurt 2005). But assume for the sake of argument that the quoted tweet reflects a belief of his.



Just as we are morally answerable for the effects of bad moral luck because we are morally required to recognize the limitations of our ability to control the outcomes of our actions, we are *epistemically* answerable for our false beliefs, even those we acquired innocently, because we are epistemically required to recognize ourselves as creatures with limited capacities to ascertain what is true. That is to say, we are epistemically answerable for our beliefs because of our *fallibility*; an omniscient mind would not need to justify its beliefs or make any effort to revise them, because by its nature everything it believed would be true. I had good grounds for believing that my mug was on my desk, since the mug that *was* there was of the same mass-produced type, but my epistemic innocence does not mean that I am not epistemically required to take responsibility for believing incorrectly; I should still update my beliefs. And even if we suppose that Trump innocently came to believe that climate change is a hoax perpetuated by a Chinese conspiracy, when we provide evidence that his belief is unjustified and false, he would still be epistemically required to update his beliefs and to issue a correction.

Our concepts, as the building-blocks of our beliefs and other attitudes, also fall under the scope of our epistemic answerability. Thus, McDowell observes that the use of concepts 'takes place under a standing obligation to reflect' about how well those concepts are serving their representational function (McDowell 1996, 12ff.). Most of the concepts that we have inherited represent the world well, but 'there is no guarantee that the world is completely within the reach of a system of concepts and conceptions as it stands at some particular moment in its historical development. Exactly not; that is why the obligation to reflect is perpetual' (40). That is to say, there is always the possibility that we might discover some new information, or that social conditions may change, or that some new range of experiences becomes possible, or that a marginal perspective comes to light, and each of these prompts critical reflection into the concepts that we use to represent and interpret what goes on. When this happens, we

may find that we must revise, replace, or retire concepts and conceptions already in our repertoire, or that we must acquire or invent concepts that we do not already have. *Critical reflection* is the central activity through which we take responsibility, epistemically speaking, for our concepts.

For the same reasons, we are sometimes also *morally* answerable for our concepts, even when our conceptual incapacities are due to bad luck. The structure of the kinds of cases with which I have been concerned, where the use or ignorance of certain concepts is morally wrongful, is the same as cases of bad moral luck more generally. Where the agent is not at fault for her wrongful conceptual incapacities, she will still be morally answerable for them because her limitations as a moral agent also provide her with reasons to take responsibility, morally speaking, for the wrongs and harms produced by her conceptual incapacities. The agent may owe an apology to those affected by her conceptual incapacities — for example, if her ignorance of the concept NON-BINARY GENDER led her to take some discriminatory or exclusionary action, then she may owe an apology to any non-binary people affected by her ignorance. But the central activity of taking moral responsibility for one's conceptual incapacities will be the same as the central activity of taking epistemic responsibility for one's conceptual incapacities: namely, to *critically reflect upon and revise one's concepts*.

Williams adumbrates this moral-epistemological activity of reflecting upon our concepts in his sketch of thick ethical concepts and their embeddedness in our cultural practices.<sup>104</sup> He develops this idea in consideration of the difficulties raised by the issue of ethical knowledge with respect to quite different cultural outlooks. One important circumstance that sets off critical reflection upon our ethical concepts is when a different way of life strikes us as what Williams calls a *real option*. In contrast to merely *notional* options, 'an outlook is a real option for a group... if they could live inside it in

---

<sup>104</sup> See Chapter 1 for the background on thick concepts.

their actual historical circumstances and retain their hold on reality, not engage in extensive self-deception, and so on' (Williams 1985, 160–61). Williams does not elaborate much further on the sort of factors that can make an option merely notional rather than real. They would seem to include epistemic or material obstacles such as having to believe a great deal of things one knows (perspective-independently) to be false (the life of an alchemist), or having to make significant and unrealistic changes to one's social situation (the life of a billionaire). They may also include complications arising from the details of the outlook under consideration, such as relying on a cultural background of practices that has died out (the life of a knight errant), has not yet come to be (the life of an interplanetary explorer), cannot come to be (the life of a Jedi), or is somehow barred to the agent (a culture open only to those born into it). Each of these obstacles requires that one make significant changes to one's way of life that one could not authentically make — attempting to live such a life in one's present circumstances would always remain make-believe, ironic, or deluded. When presented with a real option, by contrast, the other's way of life is plausible for the agent to assume, and she may take this as a reason to reflect more deeply on her own thick concepts *vis-à-vis* the alternative. For these different ways of living come with different cultural backgrounds against which different thick ethical concepts function.

For example, consider the progressive shifts of perspective that Tevye the dairyman shows in *Fiddler on the Roof* (Stein, Bock, and Harnick 1964). Tevye begins the story with relatively conservative social views, placing great importance on his Jewish cultural traditions, especially those concerning the roles of men and women, parents and children, and, in particular, the arrangement of marriages within the community. Over the course of the story, his daughters push Tevye's conception of *marriage*, and along with it, his traditional way of life, to their limits. Tevye begins with a traditional view of marriage as an economic partnership to ensure the prosperity of each household involved, where unions are properly determined by a matchmaker in

consultation with the betrothed's parents. But then, his eldest daughter, Tzeitel, spurns an arranged marriage with a rich old butcher in favour of a poor young tailor. Shortly thereafter, his second daughter, Hodel, chooses to become engaged to a young political activist, who will soon leave the village to join the resistance against the Tsar, without first asking Tevye's permission. Tzeitel's and Hodel's actions undermine Tevye's traditional perspective, supplanting the microeconomic aspects of his conception of marriage with the overriding importance of romantic love. In both cases, Tevye reflects deeply upon his conception of marriage and how it fits into his community's way of life. And in both cases, he finds that changing his conception of marriage — and with it, participating in the changing of tradition — would bring his daughters greater happiness while at the same time not threatening the continued existence of his community. He revises his conception, and approves of Tzeitel's marriage and Hodel's engagement.

But Tevye's open-mindedness has limits. When his third daughter, Chava, decides to marry a Christian, Tevye engages in reflection as before, but finds that he has at last reached the point where, as he puts it, 'If I try and bend that far, I'll break'; he has finally hit upon a merely notional option.<sup>105</sup> Accepting a marriage outside of the Jewish faith would amount to turning his back on his faith and his people — for, in his society, another major purpose of the practice of marriage is to ensure the continuation of the Jewish people through matrilineal descent. Particularly in the context of Imperial Russia, where Jews were regularly persecuted by the Christian majority, modifying the practice of marriage to allow Chava to marry outside the faith would present an

---

<sup>105</sup> Williams claims that 'Today, all confrontations between cultures must be real confrontations', given that modern life in all places is marked by a great deal of reflectiveness (Williams 1985, 163). It is not clear whether this observation applies in Tevye's case, since his is a highly traditional society and his degree of open-minded reflectiveness is portrayed as unusual for a man of his class and generation. Regardless, I am not entirely convinced that Williams is right about our present situation — can Da'esh militants rightly be said to consider my secular Western way of life to be a 'real option', or *vice versa*?

existential threat to the Jewish minority. Too much would have to change about the social context for Tevye to accept this conceptual modification. Painfully, he gives Chava an ultimatum: either she must break off her relationship, or she will be forced out of his family and community. Whatever we may think of Tevye's conclusion, his final refusal to push his reflective criticism of tradition any further illustrates how critical reflection, taken far enough, begins to break down not just individual concepts but also their cultural background.<sup>106</sup>

When Tevye's daughters confront him with non-traditional conceptions of marriage, they are taking an action similar to the moral critic in demanding that the agent justify her actions, or the epistemic critic in demanding that the agent justify her beliefs. They call on their father to take responsibility in a moral sense for his view of marriage by considering how that traditional conception may be harmful to them. Tzeitel, Hodel, and Chava will be made greatly unhappy if they are forced to marry as Tevye decides instead of how they have chosen for themselves, based on love. They are also plausibly calling on him to take responsibility in an epistemic sense by directing his attention to changing social conditions that Tevye's conception of marriage should recognize. As Tevye himself remarks, 'Love. It's a new style. On the other hand, our old ways were once new, weren't they?' In the first two cases, Tevye's reflection weighs these moral and epistemic considerations, and he concludes in favour of loosening his traditional conception of marriage to the benefit of Tzeitel and Hodel. But in Chava's case, when he is forced to weigh these considerations against a potential threat to the continued survival of his faith and culture, he is not persuaded.

Williams says little about how such critical reflection should proceed or be brought to a satisfactory close — indeed, he seems pessimistic about the prospects of

---

<sup>106</sup> Indeed, the modification and breakdown of tradition in the face of modern ways of living is the central theme of the story, ultimately symbolized by Tevye's whole community being forced out of their homes by order of the Tsar.

many of our thick ethical concepts in the face of modern reflectiveness. He suggests that philosophy might have a role to play in resolving the uncertainty these reflections may cause, but it would be a kind of vanity to assume that philosophical reflection and rational argument could handle this task alone. He also resists the idea that in every case we should aim to arrive at the *correct* set of concepts and conceptions, for he is keenly aware that when we modify our concepts, we often are also changing the cultural background against which those concepts function. That is to say, a conceptual change sometimes reflects a shift from one position of knowledge to another, by contributing to a change in the social and ethical entities that those concepts track. In the example above, Tevye moves from a traditional piece of knowledge — marriage is an economic union — to a progressive piece of knowledge — marriage is a romantic union. In changing his conception of marriage, it is hard to tell whether he is adapting to cultural change or contributing to it (probably it is both). But either way, in changing his conception of marriage, he is shifting from a perspective from which the economic conception of marriage was right, to a perspective where the romantic conception is right.

Instead of correctness, then, Williams advocates instead for a standard of *confidence*. Reaching confidence in our concepts is not just a matter of individually being convinced that one set of concepts and practices is better than another. Rather it is 'basically a social phenomenon', requiring 'social confirmation and support' of one's attitudes, concepts, practices, and judgements; and 'it is a social and psychological question what kinds of institutions, upbringing, and public discourse help to foster it' (Williams 1985, 170). We can see one way this social confirmation and support may be achieved in Tevye's story. There is some initial resistance in the community to the romantic conception of marriage, erupting in a heated argument at Tzeitel's wedding. But Tevye's spirited defence, the acquiescence of the presiding rabbi, and the general air of celebration silence the objections of nearly all. Only those who personally stand

to lose out in the changing customs — the butcher who Tzeitel rejected and the matchmaker who suggested the original arrangement — fail to feel conviction in the new conception.

Williams's remarks about confidence are suggestive, but in order to fill out the picture of how to take responsibility for our concepts, more needs to be said about the various ways in which reflection might proceed, and how we might reach confidence in our concepts. In particular, if there is a role for philosophy in this process, more needs to be said about what we can contribute and how our reflections can fit in with the cultural processes underlying conceptual change. I turn to these issues in the next chapter.

#### 4.4. Summary

In this chapter, I have developed the theoretical background for an account of how we should take responsibility for our conceptual incapacities. I started by outlining the moral notion of taking responsibility, which I identified with doing what we are morally required to do in virtue of being answerable for some action or mental reaction of ours. I then extended the account of taking responsibility to include cases of bad moral luck and penumbral agency, where despite not being not at fault for our actions or those of other agents, we are still required to take responsibility for them. I argued that we are answerable for these events, *contra* Enoch's suggestion that before we are under any obligation to take responsibility for these events we first must assume responsibility for them. I then returned to the issue of our conceptual incapacities, arguing that even if we are morally and epistemically innocent for them, we are still required to take responsibility for our conceptual incapacities. Drawing on a comparison between moral and epistemic notions of answerability, and on Williams's sketch of reflection on our ethical concepts, I suggested that critical reflection is the central activity of taking responsibility for our concepts. In the next chapter, I fill in this

sketch, describing a variety of ways we might discharge our moral-epistemic obligations to reflect upon our concepts.



## 5. Taking Responsibility for Our Concepts

In the first three chapters, I built the case that we may be culpable for our conceptual incapacities. I have shown that using certain concepts can be problematic morally, epistemically, and socially. I have also explained how conceptual ignorance and non-competence contribute to a variety of hermeneutical injustices. After defending a preferred account of moral responsibility, I argued that on this account, we can be culpable for using problematic concepts and conceptions, or for being ignorant or non-competent with regard to appropriate concepts and conceptions. In the previous chapter, I argued that we may be required — morally and epistemically — to take responsibility for things that cannot be properly attributed to our agency. With regard to our concepts, regardless of whether we are culpable for our conceptual incapacities, we are morally and epistemically required to take responsibility for them by engaging in critical reflection.

In this chapter, I explore what taking responsibility for our concepts means in greater detail. In the first section, building off the suggestion from Williams that intercultural confrontation can provoke reflection on our ethical concepts, I outline various other circumstances where we can be prompted to reflect, which can cause a loss of confidence in our concepts. In the second section, I consider a recent movement in analytic philosophy, conceptual engineering, which offers a distinctively philosophical method for subjecting our concepts to critical reflection. Even though it is often reasonable, when taking responsibility for our conceptual incapacities, to defer to expert theorists who have done some conceptual engineering, there are limitations to this approach, not least of which is that it does not address the problem that regaining conceptual confidence and effecting conceptual change is not a matter that can be settled by individual reflection or the development of philosophical theories. On the contrary, conceptual engineering must be supplemented by social processes whereby communities come to reinforce their conceptual confidence and change their concepts

by changing their practices. In the third section, I model these social processes on the classical pragmatist account of inquiry, particularly the theory of moral inquiry developed by Elizabeth Anderson. This account shows that to take responsibility for our concepts, we must take part in the social process of conceptual inquiry in good faith, though we need not always take a particularly active role.

### 5.1. Occasions to Reflect

As I argued in the previous chapter, we are answerable for our use or ignorance of various concepts, even when we are not culpable for doing so. This answerability sometimes invokes what McDowell (1996) calls a ‘standing obligation’ — a general moral-epistemic requirement — to reflect critically upon our concepts. One occasion where this requirement arises is, as Williams (1985) notes, when we are confronted with a different culture’s ethical outlook, and find that it presents a real option for us. Another way for reflection to be prompted is through feelings of alienation from some way of carving up the social world, without the presence of a ready-made alternative outlook. Sabina Lovibond describes such situations as arising when one encounters ‘other people’s talk about things for which, while not necessarily viewing them in the light of outright fictions, one would be unwilling (as it were) to take full ontological responsibility’ (Lovibond 2015, 136). For example, suppose your colleague describes a student as clever enough, but lacking a ‘first-class mind’. You might nod along with this assessment while harbouring doubts about the very notion of a first-class mind. Or you may disagree regarding your colleague’s use of the concept (that student, you believe, genuinely does have a ‘first-class’ mind), and the presence of disagreement makes you uncertain about the concept, or at least, your own conception. Or you may be ambivalent about the idea of a ‘first-class’ mind, sometimes behaving as though there were such distinctions to be made and at other times not, depending on what is convenient. Lovibond suggests that these kinds of unease regarding a concept should prompt us to think more deeply about the ways that concept carves up the social

world. For in these cases, mere ‘passive compliance’ (138) with another person’s usage is inappropriate, obscuring an attitude that is ‘ontologically sceptical, tentative, or irresponsible’ (137). These feelings of alienation from a concept are an invitation not just to take responsibility through critical reflection on your own, but also by confronting your colleague to do the same by asking for his justificatory reasons, if any, for employing that concept or conception.<sup>107</sup>

We can also be prompted to reflect upon our concepts when we observe that certain people make a point of refusing to use them. For example, consider this episode from Oscar Wilde’s testimony in his libel suit against the Marquess of Queensberry. Wilde’s complaint was that Queensberry had defamed him by calling him a sodomite in a note scrawled on a calling card. In order to undermine Wilde’s claim, the defence attorney pressed Wilde to say whether he considered a short story that glorified an abusive relationship between a priest and an altar boy to be blasphemous, hoping to catch Wilde admitting that he saw nothing immoral about sodomy or pederasty. Wilde replied several times that he disliked the story on artistic grounds before his famous retort: ‘I think it is horrible. “Blasphemous” is not a word of mine’ (Wilde v. Queensberry 1895).<sup>108</sup> Brandom cites this quip as a case of refusing to use a concept on the grounds that one does not endorse its implications (Brandom 1998, 126). On his view, these implications are spelled out in terms of the inferences one is entitled to make and the propositions one is committed to on the basis of using this concept. In the case of *BLASPHEMOUS*, the implications of this concept seem to include a commitment to the view that there exist divine beings of some kind that can be

---

<sup>107</sup> It is possible that we sometimes have obligations to voice our disagreement with the use of certain concepts. On obligations to voice disagreement regarding the truth of propositions, see Johnson (2018b, 2018a). It is plausible that these obligations generalize to other kinds of disagreement.

<sup>108</sup> Williams observes that this exchange ‘illustrates that the question of what your repertoire of thick concepts is reveals your own or your society’s ethical attitude’ (Williams 1995, 237). Williams erroneously cites the exchange as being about *obscenity*, not blasphemy. But the point is the same.

wrongfully disrespected through our actions.<sup>109</sup> Simply saying that something is or is not blasphemous does not take issue with the concept of blasphemy; a ‘not one of my words’ kind of reply rejects the concept and its associated cognitive, moral, and social implications. When we observe someone like Wilde confidently refusing to employ a concept, it may prompt us to call that concept into question by considering just what our use of that concept commits us to.

Exploring the history of a concept and the ways in which conceptions have changed over time can also prompt reflection on the conceptions we use now. This essentially describes the method of a debunking genealogy, such as Friedrich Nietzsche’s account of the origins of (post-)Christian moral concepts in resentment towards the values of ancient pagan ruling classes (Nietzsche 1998). In developing this historical narrative, Nietzsche hopes to make way for a different system of values by undermining our sense that the moral framework we have inherited is inevitable, irreplaceable, or uniquely authoritative. It is also possible to read Michel Foucault’s overall philosophical project in terms of tracing the history of various social-scientific concepts in order to unsettle their present forms. Foucault described his works as aiming ‘to create a history of the different modes by which, in our culture, human beings are made subjects’, that is to say, a history of the ways in which scientific categories, binary human categories, and our own self-conceptions restrict the ways in which we think about and behave towards ourselves and others (Foucault 2003, 126). Read this way, some of Foucault’s examples include the concepts INSANE (Foucault 1988), CRIMINAL (Foucault 1995), and PERVERT (Foucault 1990). The hope, on Foucault’s view, is that critically reflecting upon the history of these concepts will enable us to

---

<sup>109</sup> In Wilde’s case, his additional refusal to say that the story was *immoral* appears to stem from other interesting views of his to which the defence attorney made reference. For instance, in the preface to the second edition of *The Picture of Dorian Gray*, Wilde writes, ‘There is no such thing as a moral or an immoral book. Books are well written, or badly written. That is all’ (Wilde 1908, 5).

‘promote new forms of subjectivity’ (134) — or as I would put it, new concepts and conceptions — and concomitant changes to our social institutions.<sup>110</sup>

Another way to be prompted to reflect on our concepts is when someone holds us accountable for some way of thinking. Recall that on the scheme I outlined in Chapter 3, accountability is a matter of what third parties may appropriately say or do to the agent in response to her being culpable or otherwise required to take responsibility for something. The archetypal form of holding someone accountable for something is to direct some form of blame or criticism at her for a morally bad action for which she is at fault, including demands that she take responsibility for her actions. One purpose of such reproach is to make the person being blamed understand that what she did was wrong, and to correct any ignorance that may have contributed to her taking the act in question. With regard to conceptual incapacities, since it is not possible to observe the composition of other people’s propositional attitudes directly, demands that the agent take responsibility for her conceptual incapacities will often come in response to what she has *said*, where the use of certain terms rather than others suggest the use of certain concepts. Suppose you reprimand your colleague when he says, ‘She’s clever, but I don’t think she has a first-class mind’, saying something like, ‘What do you mean, “*first-class mind*?” That’s not a thing!’ What you are reprimanding him for is not merely a matter of using the wrong words, nor even primarily his unfairness to the student that his statement implies, but the use of the concept FIRST-CLASS MIND in his thinking. There is something problematic about the fact that he takes such a distinction seriously in the first place. The blame and criticism you direct at your colleague is partly to call out his error, which should prompt him to

---

<sup>110</sup> A conceptual genealogy need not be in the service of debunking a concept. Edward Craig (1999) and Williams (2002), for instance, view their genealogical projects — concerning concepts related to knowledge and truthfulness, respectively — as *vindicating* the social and theoretical roles of these concepts.

reflect upon the very notion of a ‘first-class mind’ and encourage him to reject this concept.

However, we do not always find that the agent’s statements indicate the concepts she has used so clearly; in many cases, the judgement that the agent must take responsibility for her concepts relies upon interpreting her non-verbal behaviour in light of the action-guiding aspect of the concepts we take her to be using. Consider the example of the concept SEXUAL HARASSMENT and the recent attention this and related social problems have received in public discourse, particularly through the ‘#MeToo’ movement. The movement came to public consciousness in late 2017 following a tweet from actress Alyssa Milano asking women who had experienced sexual harassment or assault to reply using the hashtag ‘#MeToo’ to ‘give people a sense of the magnitude of the problem’ (@Alyssa\_Milano 2017).<sup>111</sup> The response to Milano’s tweet precipitated a series of allegations against men in powerful positions, particularly in politics, radio, film, and TV. As essayists responding to the movement have noted, these allegations and personal stories of sexual harassment and assault are part of a larger cultural conversation over what sort of behaviour should be seen as sexual harassment and sexual assault, or mere awkwardness or rudeness.<sup>112</sup> As philosopher Emily McWilliams remarks in an interview for the *Examining Ethics* podcast, ‘We have a concept of sexual harassment but we don’t have the *right* concept, or at least, we don’t have the right concept across all social locations’ (Price and Wisheart 2018, 14:55–15:05). Compare these remarks by columnist Hadley Freeman:

‘Ask a man for his worst date story, and he’ll tell you something that will make you laugh’, a female friend said to me a few years ago. ‘Ask a woman, and she’ll tell you a story that will make you want to change the locks’... [It] wasn’t until recently that I realised the man and the woman in these scenarios could be talking about the exact same date. A few weeks ago, as by now everyone knows,

---

<sup>111</sup> The quote is from an image shared with Milano’s tweet. An earlier movement of the same name was started by activist Tarana Burke in 2006 as an awareness and empowerment campaign for survivors of sexual violence (Burke n.d.).

<sup>112</sup> See, among others, Hamblin (2018), Filipovic (2018), and Friedersdorf (2018).

a woman using the pseudonym Grace published an account of her date with US comedian Aziz Ansari, a date that he thought was ‘fun’ with ‘completely consensual sexual activity’, and that she described as ‘the worst night of her life’, riddled with ‘assault’. (Freeman 2018)<sup>113</sup>

The Ansari story shows how two people may have radically different interpretations of the same shared experience. The behaviour of the parties involved and the use of different terms to describe the event — ‘consensual’ vs. ‘assault’ — suggests that part of the difference lies in their differing use of concepts and conceptions. In the terms I have been using, the collective hermeneutical resource contains the concept of sexual harassment, but there are various competing conceptions circulating amongst different social groups, some of which contain misunderstandings that amount to non-competence with the concept. One important component of the #MeToo movement (though perhaps not the most visible part) is to prompt reflection on our ability to correctly identify and respond to sexually harassing behaviour. This demand may arise in the course of reproaching the conceptually non-competent agent for his problematic behaviour stemming from just such a mistaken conception of sexual harassment.

One may also be prompted to reflect on one’s concepts when one experiences a hermeneutical injustice oneself. Staying with the example of sexual harassment, recall that after repeatedly encountering difficulties in expressing her experience intelligibly, Carmita Wood discussed the issue with a group of feminist activists. Once they realized that they had all experienced similar abuses and had all similarly found it difficult to communicate about their experiences, as a group they reflected on the nature of this social phenomenon and what they should call it in order to organize against it. Concepts that others had used to interpret the men’s behaviour — perhaps including notions such as flirtatious, playful, or unprofessional, but never *wrongful* — were clearly inadequate. According to Brownmiller (1990), the group tried various options

---

<sup>113</sup> The initial reporting of the Ansari story has been criticized for journalistic and editorial failings, but the case is still illustrative. It also, regrettably, is a more common feature of dating than many would like to admit.

before settling on 'sexual harassment'. Their shared experience of sexual harassment *avant la lettre* prompted Wood and her fellow activists to reflect, producing a new concept that captured the nature of their experience, enabling them to speak out against this prevalent moral problem and to overcome the moral-epistemic wrongs of hermeneutical injustice.

Finally, a proper awareness of the fallibility of the conceptual repertoire that we have acquired, particularly given the presence of hermeneutical marginalization, should prompt reflection on our concepts and conceptions even before we find ourselves being held accountable by others or subject to a hermeneutical injustice ourselves. Ideally, this heightened awareness of the possibility that one's concepts are inadequate will manifest as an ethical-epistemic virtue that Fricker calls *hermeneutical justice*. She defines this virtue as a corrective to the lack of communicative intelligibility experienced by subjects of hermeneutical injustice when they attempt to convey their experience to a hearer who lacks the necessary concepts or conceptions. Thus, the hermeneutically just hearer possesses

an alertness or sensitivity to the possibility that the difficulty one's interlocutor is having as she tries to render something communicatively intelligible is due not to its being a nonsense or her being a fool, but rather to some sort of gap in collective hermeneutical resources. (Fricker 2007, 169)

Minimally, when the virtuous hearer encounters a speaker whose testimony seems barely intelligible, and whom she recognizes is likely to be hermeneutically marginalized, she will give the speaker the benefit of the doubt and assume that even though she does not understand what has been said, the speaker is not spouting nonsense. Making this charitable move may prompt the hermeneutically just hearer to reflect upon her own concepts and conceptions, opening up a space for the marginalized speaker's experience to find some intelligible formulation. In favourable conditions, the hearer and speaker may be able to carry on an extended conversation aimed at producing or sharing the required concepts and conceptions so that both



parties are able to make sense of the marginalized subject's experience. The virtuous hearer need not accept the marginalized speaker's alternative hermeneutical tools, for as I argue below, some conceptual innovations or revisions should be rejected.<sup>114</sup> But in any case, the virtuous hearer will reflect critically upon the concepts she uses to interpret the marginalized speaker's testimony, at least leaving open the possibility that her own concepts and conceptions do not enable her to understand.

Consider the following example, drawn from Marilyn Frye's reflections on white feminism and racism. Frye describes her experience as a member of a feminist consciousness-raising group created to discuss and address racism, which initially was to be open only to white women. Frye and her colleagues consulted with some women of colour, and were surprised when these women voiced concern that an all-white anti-racism group would itself be racist. Not fully understanding their reasoning, the white feminists invited the women of colour to attend the initial meetings as well. At a later meeting, one of these critics angrily attacked the idea that the group had at any point thought they could achieve anti-racist goals working only with white women. Frye and others responded:

We said we never meant this few weeks of this particular kind of work to be all we ever did and told her we had decided at the beginning to organize a group open to all women shortly after our series of white women's meetings came to a close... [We] could hardly have said anything less satisfying to our critic. She exploded with rage: '*You* decided!' Yes. We consulted the opinions of some women of color, but still, we decided. 'Isn't that what we are supposed to do?' we said to ourselves, 'Take responsibility, decide what to do, and do something?' She seemed to be enraged by our making decisions, by our acting, by our doing anything. It seemed like doing nothing would be racist and whatever we did would be racist just because we did it. We began to lose hope; we felt bewildered and trapped. It seemed that what our critic was saying must be right; but what she was saying didn't seem to make any sense.

She seemed crazy to me.

That stopped me.

I paused and touched and weighed that seeming. It was familiar. I know

---

<sup>114</sup> *Pace* Komarine Romdenh-Romluc, who worries that hermeneutical justice requires that we accept noxious alternative social views, such as those of white nationalists (Romdenh-Romluc 2016).

it as deceptive, defensive. I know it from both sides; I have been thought crazy by others too righteous, too timid and too defended to grasp the enormity of our difference and the significance of their offenses. I backed off. (Frye 1983, 111–12)

Frye presents this episode as a moment of reckoning for her conception of racism.

Before, she had considered racism only as the kind of transparently hateful acts and opinions of explicit racists. But her discussions with feminist women of colour revealed another conception of racism, which captures systemic social structures, subtle patterns of behaviour, and the social power and privilege possessed by white people, of which one may be the beneficiary *without* holding racist beliefs. Because the latter conception was absent from her repertoire, the criticism Frye received seemed unintelligible to her. The women of colour were insisting that Frye take responsibility for the concepts she used to interpret her own behaviour, particularly by taking up the conception of racism as social-structural that anti-racist activists had already developed. In questioning her initial judgement that her critic seemed crazy, and reflecting critically upon the individualist conception of racism she had been using *vis-à-vis* that of anti-racist activism, Frye took responsibility via the hermeneutically just action of reflecting in order to correct her own conceptual ignorance.

Once reflection has been launched, the trouble is how to return to conviction in our concepts. As Williams observes, resisting the call to reflect by clinging to tradition is no longer a reasonable response, because it shirks the responsibilities imposed on us by conscientious modern life: 'there is no route back from reflectiveness... measures would have to be taken to stop people raising questions that are, by now, there to be raised' (Williams 1985, 163–64). Yet to follow through on the requirement to engage in critical reflection can leave us in an unsatisfying place:

reflection might destroy knowledge, because thick ethical concepts that were used in a less reflective state might be driven from use by reflection, while the more abstract and general ethical thoughts that would probably take their place would not satisfy the conditions of propositional knowledge. (167)

The same thing can happen with regard to our social concepts: reflection can undermine our conviction that these give us the right way to make distinctions in the social world, as in the examples of the concepts *FIRST-CLASS MIND* or *INSANE*. Once we start to take responsibility for our concepts, we can lose the social and moral knowledge enabled by those concepts, and the practices and institutions organized around them.<sup>115</sup>

Williams considers and rejects several ways to carry on once reflection has undermined the conviction we once had in our concepts. The reassertion of tradition on the basis of *certainty* is not an acceptable response, because part of what is in question are the very grounds of that certainty. Another option might be to give up the search for conviction, and revel in the ‘refined indecision’ that one can reach by reflecting without end (Williams 1985, 169). But this is intolerable; not only is it practically useless, giving us no stable concepts around which to coordinate our social lives, it still shirks our rational responsibilities, just at a later stage: the postmodernist romanticisation of uncertainty is to opt for whim and caprice without justification. Another source of ethical conviction might be ‘a *decision*, to adopt certain moral principles or to live in one way rather than another’ (169). But this model cannot be right, either, because making that decision must be based on a conviction one already

---

<sup>115</sup> For Williams, ethical knowledge simply requires knowing how to apply one’s thick ethical concepts within one’s cultural outlook, and not a matter of grasping some objective moral facts. But we need not accept his form of relativism to adopt the general insight his view offers. The basic idea of a reflective cycle — where some experience prompts us to reflect on our concepts, which causes us to lose confidence in our concepts, followed by either return to confidence in the same or different concepts and conceptions or the rejection of certain concepts altogether — is compatible with other meta-ethical stances. A moral realist could interpret the reflective cycle as aiming at uncovering ethical truth, for objective knowledge comes with its own kind of confidence. And an anti-realist could translate Williams’s cognitive terminology into terms referencing their preferred conative entities. I will not take a stand on the issue of ethical knowledge here; I simply flag it to set aside worries about how, below, I tend to put things in terms of removing ignorance and making progress. Even though, as I argue below, there is good reason to think (and not just from a parochial perspective) that we can make genuine improvements to our concepts, whether this improvement is a matter of uncovering objective truth is not my concern.

has — or else the arbitrariness of that decision immediately reopens the reflective questions one sought to silence.

As noted in Chapter 4, Williams introduces the notion of *ethical confidence* to provide the needed sense of conviction in our concepts. His sense of confidence is not merely that of an individual psychological state. Rather, it is ‘basically a social phenomenon’, requiring ‘social confirmation and support’ of one’s attitudes, concepts, practices, and judgements (Williams 1985, 170). We start in a position of confidence in our concepts, which enable us to have social and ethical knowledge based on the distinctions they make, and around which various cultural practices and social institutions are organized. But once we are prompted to reflect, if we follow through on the moral-epistemic requirement to do so, we can lose our confidence in those concepts. Whether we have an alternative perspective in view or not, once we begin to question the distinctions we make in thought, we begin to question the social background against which the relevant concepts find their descriptive and evaluative point. When these reflective worries are shared by a segment of the community, the social support that we give and receive for continuing to use those concepts is undermined. So we need to find some way of returning to confidence, or else we risk losing much of our social and ethical knowledge. We might return back to the concepts that we started with — indeed, for most of us there is a strong tendency in this direction. Or we may reach confidence in some different concepts or conceptions. But Williams says very little on the nature of the process by which a return to confidence can be effected.

In this section, I have outlined several ways in which we may be prompted to reflect upon our concepts, which can undermine our confidence in them: intercultural interactions that reveal conceptual differences, feelings of alienation from the use of certain concepts, explicit refusals to use certain concepts, genealogical accounts of the history of our concepts and conceptions, being held accountable for utterances or

actions that express the use of some concept or conception, experiencing a hermeneutical injustice, and exercising the virtue of hermeneutical justice. These occasions prompt us to take responsibility for our conceptual incapacities through critical reflection. While several of these cases suggested some ways in which we might return to confidence in different concepts or conceptions, this process requires further explication. In the next section, I outline a recent philosophical approach to critiquing our concepts that provides a plausible but, I argue, incomplete answer.

## 5.2. Philosophical Reflection

Williams can seem sceptical of the possibility of judging that one concept is *better* than another, which is crucial if we want to make sense of the idea that taking responsibility for our concepts through reflection can lead to moral, epistemic, and social improvement. This is an aspect of his more general scepticism about the posture that ethical theory tends to take, of offering a judgement of our moral systems from some external vantage point:

Any such picture makes in some degree a Platonic assumption that the reflective agent as theorist can make himself independent from the life and character he is examining. The belief that you can look critically at all your dispositions from the outside, from the point of view of the universe, assumes you could understand your own and other people's dispositions from that point of view without tacitly taking for granted a picture of the world more locally familiar that would not be available from there. (Williams 1985, 110)

On his view, this kind of external perspective is not possible in ethics, because once you have stepped back from your ethical concepts, there is no evaluative standard by which to judge that one outlook is better, ethically speaking, than another. But this sceptical point applies not to individual concepts, but to entire *schemes* of ethical concepts. Thus, Williams later remarks that his form of ethical relativism applies only 'to fairly large-scale systems or bodies of beliefs and attitudes', that is, to 'outlooks rather than particular practices' or particular concepts (162). We can remain sceptical of the possibility of judging between radically different outlooks while still maintaining a

place for reflective revision of particular concepts or closely related clusters thereof.

The picture of reflection I am taking from Williams does not force us to withhold judgement on any particular concept or cluster of closely related concepts, which are the subject of the kind of critical reflection that constitutes taking responsibility for one's concepts.

If reflection on our concepts cannot take place from the 'view from nowhere',<sup>116</sup> we are left with a model of reflection that has to take place from *within* an outlook made up of a relatively stable system of concepts and beliefs. McDowell, also critical of theory that tries to take an external perspective on ethics (or as he puts it, a 'sideways-on view' (McDowell 1996, 1998a)), describes this kind of reflection by reference to Otto Neurath's famous image of a mariner on the open sea who must make repairs to his ship without being able to hoist the boat into drydock to make a complete overhaul (Neurath 1973). We have to hold enough of our concepts and beliefs stable in order to judge the merits of particular concepts and to make changes as appropriate. However, as McDowell points out,

this does not mean that reflection cannot be radical. One can find oneself called on to jettison parts of one's inherited ways of thinking; and, though this is harder to place in Neurath's image, weaknesses that reflection discloses in inherited ways of thinking can dictate the formation of new concepts and conceptions... the essential thing is that one can reflect only from the midst of the way of thinking one is reflecting about. (McDowell 1996, 81)<sup>117</sup>

Making piecemeal adjustments in this way, we can isolate our conceptual incapacities and improve our concepts and conceptions.

How do we proceed in such reflection? Individual reflection may proceed in a way similar to Tevye's thoughts about marriage, as described in Chapter 4. By

---

<sup>116</sup> This turn of phrase is Nagel's (Nagel 1989).

<sup>117</sup> McDowell (1996) generalizes this point, using ethical concepts and judgements as a model for empirical knowledge, but we can take the point about reflection on our social and ethical concepts without bringing along his more controversial epistemological commitments.

considering the different practical consequences of adopting one conception over another, he comes to accept a view of marriage as a romantic rather than merely economic union, though he stops short of permitting marriage outside of his religious community. However, not all of us are as conscientious and observant as Tevye, and few of our defective concepts are revealed to us quite as dramatically as his conception of marriage is thrown into doubt by his daughters' actions. Thus, Mary Midgley argues that when our concepts and conceptions are in need of revision, we would do well to turn to professional philosophers for guidance. She makes this argument by analogy to plumbing. The vast networks of pipes, sewers, pumps, and treatment plants, on the one hand, and the systems of concepts and conceptions that we have inherited, on the other, each developed slowly over centuries, never had a complete top-down plan, are constantly being altered to suit changing needs, are too complex and integral to life as we know it to replace entirely from the ground up, and sometimes go wrong and need repair. But when it comes to making these repairs,

About plumbing, everybody accepts the need for trained specialists. About philosophy, many people... not only doubt the need, they are often sceptical about whether the underlying system even exists at all. It is much more deeply hidden. When the concepts we are living by work badly, they don't usually drip audibly through the ceiling or swamp the kitchen floor. They just quietly distort and obstruct our thinking... We may indeed complain that life is going badly — that our actions and relationships are not turning out as we intend. But it can be very hard to see why this is happening, or what to do about it... When things go badly, however, we have to do this... That need to readjust our concepts is the need that philosophy exists to satisfy. (Midgley 2006, 1–2)

Midgley notes that non-specialists may succeed in analysing, critiquing, and improving their concepts. But the background knowledge, skills, and resources available to professional philosophers are often required just to discover that there are problems with some concept in the first place. Professional philosophers are thus often in a better position than most to work out the intricacies of how some concept fits into our lives, in just what respects it is confused or damaging, and what sorts of remedy might be effective.

Others have described one of philosophy's central roles as the critique and improvement of our concepts. W. V. O. Quine, for example, characterizes Rudolph Carnap's method of 'explication' as 'not merely to paraphrase the definiendum into an outright synonym, but actually to improve upon the definiendum by refining or supplementing its meaning' (Quine 1953, 25). The idea is that philosophical analysis aims not just at describing the concepts we already employ but also to suggest improved versions of those concepts where this will contribute to clarity of thought or to precision or accuracy in theoretical inquiry. Peter Carruthers, in what he calls a manifesto for 'conceptual pragmatism', argues that where we agree that a particular concept is in use, we should 'ask whether it is this concept which we ought to employ, given the nature of our purposes', and 'where there is uncertainty or dispute over the concept which we in fact employ, our difficulties may be either resolved or dissolved by investigating which (if not [all]) of the candidate concepts are appropriate to our purposes' (Carruthers 1987, 218).

This approach has attracted renewed attention, under the labels 'conceptual engineering' (Cappelen 2018), 'conceptual ethics' (Burgess and Plunkett 2013a, 2013b), or 'ameliorative inquiry' (Haslanger 2012c, 2012d). Sally Haslanger distinguishes this method from two other philosophical methods aimed at critical scrutiny of our concepts. What she calls *conceptual* projects analyse, from armchair intuitions, what we take a particular concept to mean; this kind of analysis identifies what Haslanger calls our 'manifest' concepts. For example, an armchair analysis of the concept COOL might lead to the conclusion that a cool person is calm, confident, stylish, and powerful. *Descriptive* projects make use of empirical data, possibly taken from natural-scientific, social-scientific, or experimental-philosophical studies, to determine which properties or kinds (natural or social), if any, a particular concept actually tracks; this kind of project reveals our 'operative' concepts. It may be, for instance, that COOL actually



tracks conformity to prevailing norms for one's social group.<sup>118</sup> *Ameliorative* projects, by contrast, ask what purpose a particular concept serves or what the practical upshot of using that concept is, whether that purpose is commendable or whether that practical result is acceptable, and whether a different concept or conception might serve our ends better; this kind of project identifies what Haslanger calls our 'target' concepts. For instance, we might decide that the purpose we would like COOL to serve is to acclaim people who buck the constraints of convention with regard to artistic expression or personal style, then provide a candidate analysis of this revisionist conception of COOL. As Haslanger puts it, 'The responsibility is ours to define them [i.e., our concepts] for our purposes' (Haslanger 2012a, 224). All three methods are useful in our critical reflections on our concepts, but only conceptual engineering makes specific recommendations for improvement.

For example, Robin Dembroff employs the ameliorative method in an essay on the concept of sexual orientation. The received view of sexual orientation that Dembroff criticizes is defined by reference both to the subject's own sex or gender, and to the sex or gender of the people to whom the subject is sexually attracted. A man and a woman who are both attracted to men thus have *different* sexual orientations. In setting up the desiderata for an ameliorated conception, Dembroff claims that one thing we should aim at is for our conception of sexual orientation to be 'conducive for establishing legal and social protections for persons who have queer sexual orientations' (Dembroff 2016, 5). They then observe that a conception of sexual orientation that ignores the subject's *own* sex or gender is instrumental in making such legal claims. For example, if a man and a woman both wish to marry a man, a law that

---

<sup>118</sup> This is plausibly suggested by Wayne Martino's (1999) study of adolescent boys at an Australian Catholic secondary school. These young men considered football players who conform to masculine norms of aggression, athleticism, unstudiousness, and masculinity to be the 'cool boys', while denigrating others as 'poofs' (for not being aggressive or masculine enough) or 'squids' (for being unathletic and studious).

allows a marriage licence to be issued only to the different-sex couple can be construed as a case of *sex- or gender-based* discrimination. These forms of discrimination are uncontroversial and well-established in existing laws and legal precedent (unlike, perhaps *sexual orientation-based* discrimination) and thus difficult to dispute in court and unlikely to be deferred to legislators to resolve (20).<sup>119</sup> Dembroff thus presents an ameliorated analysis of SEXUAL ORIENTATION as follows:

A person S's sexual orientation is grounded in S's dispositions to engage in sexual behaviors under the ordinary condition(s) for these dispositions, and which sexual orientation S has is grounded in what sex(es) and gender(s) of persons S is disposed to sexually engage under these conditions. (18)

Dembroff's ameliorated conception makes reference only to the sex or gender of the people to whom the subject is attracted, and not to the subject's own. A man and a woman who are attracted to men would thus have the *same* sexual orientation, and discrimination on the basis of whom they are attracted to would fall under established categories of sex- or gender-based discrimination.

It is tempting to offer conceptual engineering as the method we should follow for taking responsibility, both epistemically and morally, for our conceptual incapacities. It provides a philosophical procedure to follow in reflecting critically upon our concepts, and allows for moral, political, and theoretical goals to guide our decision as to which concepts and conceptions to use. Of course, not everyone is in a position to engage in precisely this kind of reflection, either because of a lack of philosophical training or inadequate time or resources to do so. Thus, in a remark that echoes Midgley's plumbing analogy, Haslanger argues that 'in the social domain we should rely on social theorists, including feminist and antiracist theorists, to help explicate the meanings of our terms' (Haslanger 2012b, 15). In her view, expert theorists are in a

---

<sup>119</sup> Dembroff cites a comment by U.S. Supreme Court Justice John Roberts in support of this argument: 'I'm not sure it's necessary to get into sexual orientation to resolve this case... I mean, if Sue loves Joe and Tom loves Joe, Sue can marry him and Tom can't. And the difference is based on their different sex. Why isn't that a straightforward question of sexual discrimination?' (quoted in Liptak 2015).

better position to work out what concepts we are actually using and how to improve them. If we accept this picture of professional philosophers as being in a unique position to offer these sorts of conceptual insights, it is reasonable to turn to theorists and philosophers when our concepts break down, just as we rely on the expertise of plumbers when our faucets break down, or indeed, just as we rely on medical experts when our health breaks down.

But we should be careful not to make this reliance on philosophers and theorists too strong. There are at least three reasons why we should not lay all of our burdens at the feet of the conceptual engineer. One reason for this is practical: theorists are not always talking about the concepts that most need engineering work, so it may be that we cannot rely on their current work to guide us. An all-too-common reason for this discrepancy is hermeneutical marginalization. The social exclusion and marginalization that gives rise to hermeneutical injustice runs wide and deep in society, and academe is not immune, so turning to theorists will not always provide a stopper for hermeneutical gaps. Recall the example of Wendy Sanford, who experienced post-partum depression in the 1960s, when this condition was poorly understood by the medical community, much less her own upper-class conservative social group. Had she turned to philosophers or medical theorists at the time, Sanford would not have found much by way of assistance in her attempts to understand her experience. Instead, she turned to a group of feminist activists, who were not themselves expert theorists or, for the most part, medical professionals. Because some social issues lack attention in academic discourse, whether because of a structural bias against them or because social issues are moving too quickly for academics to mount a timely response, taking responsibility for our concepts cannot simply be a matter of deferring to theorists.

Another reason not to simply defer to philosophers and theorists is epistemological: with regard to many contentious concepts, there is often a great deal

of disagreement among the experts, so turning to conceptual engineers will give us an array of different options without any definitive guidance as to which to adopt.<sup>120</sup> For example, there is considerable disagreement among feminist philosophers and theorists regarding the concept WOMAN. Rejecting the manifest concept of a woman as a person who has female sexual and reproductive organs, Haslanger's ameliorative analysis identifies women as members of a social class who are *socially subordinated* by virtue of being perceived to have such organs (Haslanger 2012a). She is led to this conclusion on the basis of the assumption that feminist theory should aim to eliminate inequalities between men and women; she takes the most direct approach to this end by making the distinctive form of oppression suffered by women intrinsic to that very social kind. Katharine Jenkins counters that while there is value in identifying this kind of social class, other considerations are more important with regard to who falls under the concept WOMAN (Jenkins 2016). Haslanger's analysis makes no reference to individuals' *identity* as women, which leads to the exclusion of any trans women who do not appear to be cis (i.e., not to be trans). Jenkins argues that validating the gender identities of trans women is a more important goal for a contemporary feminist conception of woman, so she develops an alternative ameliorative analysis of the concept that concentrates on the idea of how one's 'internal map' guides one through the social roles associated with being a woman. The dispute between Haslanger and Jenkins reflects a broader and highly contentious disagreement among feminist activists and theorists concerning the nature of gender.<sup>121</sup> This dispute and others like it can be difficult and intimidating for non-specialists to navigate. Where the experts disagree, merely deferring to their conclusions cannot constitute the whole of taking responsibility for our concepts, because we will be still lack a satisfactory answer.

---

<sup>120</sup> This is a variation on Alvin Goldman's 'novice/2-experts problem' (Goldman 2001).

<sup>121</sup> For a level-headed summary of the often heated disagreement over trans inclusion within feminism, see Goldberg (2014).

A third reason not to simply defer to conceptual engineers is ethical. In the previous chapter I described critical reflection as the central way in which we may take responsibility for our concepts; regardless of whether we are culpable for our conceptual incapacities, we will be morally and epistemically required to take responsibility in some way. We might think that there is something significant about requiring the person taking responsibility to do some reflective work *herself*. This requirement could be motivated by analogy to another action that is commonly required as part of taking responsibility: apologizing. If you owe someone an apology, it does not seem you have taken responsibility (at least, not fully) if you have a friend apologize on your behalf. Your connection to the event for which an apology is owed makes it the case that *you* owe the apology, and cannot ordinarily defer this answerability-obligation to someone else. Similarly, there may be cases where merely relying upon expert theorists without doing reflective work oneself may amount to deferring rather than taking up the requirement to take responsibility for one's conceptual incapacities. Something like this may be behind the view that it is not the duty of people in marginalized groups to educate people from relatively dominant groups, as Audre Lorde remarks:

Whenever the need for some pretense of communication arises, those who profit from our oppression call upon us to share our knowledge with them. In other words, [they assume] it is the responsibility of the oppressed to teach their oppressors their mistakes... The oppressors maintain their position and avoid responsibility for their own actions' (Lorde 2007, 114–15).

It is possible, then, that more is morally required of us than merely asking what the experts have to say about some disputed concept, perhaps especially when the experts are members of a marginalized group.

But there is a deeper philosophical reason to worry that we cannot simply delegate the requirement to reflect to conceptual engineers. The problem lies in the nature of conceptual confidence. Recall that Williams describes confidence in our ethical concepts as a social phenomenon. Williams is sceptical that philosophy or

theory alone could produce such confidence, though he acknowledges that philosophical argument has an important role to play. Herman Cappelen presents a similar but more pessimistic argument in his theory of conceptual engineering. If the aim of conceptual engineering is to change the meaning of representational devices such as words and concepts, we run into the problem that the processes by which meanings change 'are for the most part inscrutable, and we lack control of them' (Cappelen 2018, 72, boldface removed). This claim is a consequence of the semantic externalism that Cappelen employs; on his account, the mechanisms by which the extensions and intensions of representational devices change are too complex for us to understand in any detail, and no one, not even the best equipped theorist, has much control or influence over them. We might worry, then, just how important the conceptual engineer really is in our efforts to take responsibility for our conceptual incapacities, if philosophical reflection seems unlikely to produce the required confidence or meaning change.

A more modest role for the conceptual engineer would be not to effect a change in our concepts, but instead to reveal, by clearing up some confusion, what our concepts *actually* are. Haslanger herself makes this suggestion. Also drawing on semantic externalism, she argues that the conceptual engineer may demonstrate that her 'improved' analysis of some concept 'shows us that our assumptions about what we mean are false, given our practice. This is not to propose a new meaning, but to reveal an existing one' (Haslanger 2012d, 398). Whatever social processes lead to a change in our practices and meanings, this kind of inquiry aims at making these changes explicit by debunking our unreflective intuitions, not at directly changing those practices and meanings. Sometimes, the difference between our manifest and operative concepts will be one that has persisted, mostly unnoticed, for a long time; in other cases recent social developments will have changed one of our operative concepts while the corresponding manifest concept remains tied to a previous social

condition. Haslanger uses the example of PARENT: the manifest concept is something like 'immediate biological progenitor', but our operative concept is something closer to 'primary caregiver'. It is an empirical question whether this divergence is recent or ancient, but the long history of 'found family' narratives and ubiquity of adoption practices suggests that the caregiver sense is older than we might initially think.

Any account of this kind, that purports to show that we have been mistaken about our concepts, would have to show not just that it is correct; such an account would also have to explain why we had the mistaken conception for so long. In the case of PARENT, the first task could be done by noting that many people who are not immediate biological progenitors of the cared-for individual are treated (socially, legally, morally) as though they are. Examples include appointed guardians, adoptive parents, step-parents, same-sex spouses, and partners in polyamorous relationships. As to the second task, to explain why the operative concept has gone mostly unnoticed it is enough to observe that most primary caregivers are in fact the immediate biological primogenitor of the cared-for individual, and that much of our society is structured around the assumption that this will be so, which produces a default presumption that parents *are* immediate biological primogenitors. Once the discrepancy between the manifest concept and the operative concept has been revealed, suggestions can be made for how to improve our practices so that they are more consistent with what our concepts actually are. For example, laws about parental rights tend to explicitly assume that the primary caregivers of a child will be an opposite-sex couple, leading to difficulties when same-sex couples or parents in a more complex family structure attempt to access government programmes concerning their

child. A recognition that biological primogeniture is not necessary for parenthood would enable legal and social change to address this kind of problem.<sup>122</sup>

But as Cappelen observes, the sort of project described above is essentially a reactive kind of theorizing, and not the proactive stance aiming to effect conceptual change that ameliorative inquiry aims at: ‘What we thought was amelioration (i.e., improvement of our representational devices) turns out to be revelation of something that was there all along. This means that Haslanger effectively gives up on the ameliorative project’ (Cappelen 2018, 80). Cappelen goes on to argue that just because we do not really understand how to change our concepts and probably lack control over these processes anyhow, this is no reason not to make the effort to ameliorate. The challenge facing the conceptual engineer is just a particular instance of a more general kind of predicament we face whenever we try to change things for the better:

think about trying to make a positive change to a person’s life (say, that of a child you have responsibility for). For the most part, we understand very little of how such changes can be achieved and what we do know tells us that we have very little control. More often than not, what we do will either have no effect, or will have some effect other than what we intended. Nonetheless, we keep trying; there’s a sense in which we can’t give up. (75)

Similarly, Williams argues that while we cannot (and should not) make *certain* that our values and ethical concepts will be taken up by future generations, we should at least ensure that they are able to engage in free inquiry regarding their own values and concepts. And, ‘To our immediate successors, our children at least, we have reason to try to transmit more: it is a mark of our having ethical values that we aim to reproduce them. But this does not affect very determinately what remoter generations will hold’ (Williams 1985, 173). The role of the theorist, philosopher, or conceptual engineer, then, is to make well-reasoned *proposals* for what our concepts and conceptions *should*

---

<sup>122</sup> A recent court ruling in Newfoundland and Labrador has made a step in this direction by recognizing three adults in a polyamorous relationship as the legal parents of their child (MacDonald 2018).



be. With some further effort, and some luck, the best concepts may win the day — but there are no guarantees.

Conceptual engineering in a sense is always either too late — our concepts have already changed and the conceptual engineer simply alerts us to this shift — or too early — the conceptual engineer makes proposals and cannot directly change our concepts. So while philosophical reflection is a useful tool for taking responsibility for our concepts, and though turning to philosophers and theorists can be helpful to our own efforts in this regard, it cannot be the whole story. So what else is needed? It is suggestive that both Williams and Cappelen point to social processes as part of the means by which confidence is secured and through which meaning shifts. These processes include changing patterns of concept use and concomitant changes in social life. While philosophical reflection is a part of these social processes, it is not the only, or often even the central contributor. To get a better understanding of how to take responsibility for our concepts, we have to get a better view of what these social processes are, and how our actions fit in. In the next section, I argue that the pragmatist understanding of inquiry provides a plausible model of these processes.

### 5.3. Conceptual Inquiry

The cycle that follows from Williams's picture of ethical concepts — an experience that prompts reflection, a loss of confidence, and a return to confidence — can be interpreted as a particular form of *inquiry*, as understood by the classical American pragmatists Charles Peirce and John Dewey. On their accounts, we engage in inquiry to settle our beliefs and habits when doubt or loss of conviction make us uncertain how to act. In order 'to discover specific connections between something which we do and the consequences which result', we deliberate on a potential course of action, then test it in practice (Dewey 1980, 152). Successful experiments of this kind contribute to 'fixing' our beliefs and eliminating doubt, in turn laying down new habits of thought and action. In contrast to other modes of returning to conviction — such as tenaciously

reasserting one's beliefs, deferring uncritically to authority, or deducing from *a priori* principles — Peirce emphasizes that this method of inquiry is the only approach that 'presents any distinction of a right and a wrong way' (Peirce 1992b, 121). This is because requiring a belief or habit to be tested in practice enables us to see whether, 'if acted on, it will carry us to the point we aim at and not astray' (123). Dewey also rejects solitary armchair reflection as the standard by which doubt should be resolved: 'situations that are disturbed and troubled, confused or obscure, cannot be straightened out, cleared up and put in order by manipulation of our personal states of mind. The attempt to settle them by such manipulations involves what psychiatrists call "withdrawal from reality"' (Dewey 1986, 109).

While the pragmatists initially presented their model of inquiry as a theory of the scientific method, they went on to apply it generally. As Peirce remarks, 'Everybody uses the scientific method about a great many things, and only ceases to use it when he does not know how to apply it' (Peirce 1992b, 120). Of interest for present purposes is the application of this method of inquiry to social and moral issues. According to James Feibleman's reconstruction of Peirce's ethics from scattered remarks throughout his corpus, the experimental method of inquiry plays a role in determining which general ethical principles are acceptable: 'in the study of pure ethics, conduct plays the role of experimentation, and its allowance or rejection of ethical hypotheses is carefully noted' (Feibleman 1943, 103). Similarly, Dewey argues that

[Moral] principles exist as hypotheses with which to experiment... There is a long record of past experimentation in conduct, and there are cumulative verifications which give many principles a well earned prestige. Lightly to disregard them is the height of foolishness. But social situations alter; and it is also foolish not to observe how old principles actually work under new conditions, and not to modify them so that they will be more effectual instruments in judging new cases. (Dewey 1983, 165)

Dewey extends this social and moral application of inquiry to serve as the basis for his social, educational, and political theory; as James Bohman summarizes, 'The task of Dewey's moral theory is thus to provide an account of the practical basis for social

change that will not only make democracy possible but also a continually improving practice' (Bohman 2010, 189).

Importantly, in order to ensure that the principles to be considered and the conditions under which they are tested are free (as far as possible) from bias, inquiry must go on in as inclusive a community as possible. For Peirce, the notion of diversity invoked is rather extreme:

logicality inexorably requires that our interests shall *not* be limited. They must not stop at our own fate, but must embrace the whole community. This community, again, must not be limited, but must extend to all races of beings with whom we can come into immediate or mediate intellectual relation. (Peirce 1992a, 149)

Dewey similarly emphasizes the importance of a diverse community of inquirers whose interests must be broader than their own immediate concerns. In describing the democratic ideal, Dewey (with James Hayden Tufts) writes:

democracy signifies, on one side, that every individual is to share in the duties and rights belonging to control of social affairs, and, on the other side, that social arrangements are to eliminate those external arrangements of status, birth, wealth, sex, etc., which restrict the opportunity of each individual for full development of himself. (Dewey and Tufts 1985, 348–49)

The point is not merely to eliminate inequality in the law, but also to break down barriers to communication between different groups so they can work together on social problems: 'this specification cannot be fulfilled except when different groups interact flexibly and fully in connection with other groups' (Dewey 1984, 328). As Elizabeth Anderson summarizes, on Dewey's account this open communication between people from diverse backgrounds allows them to pool their experiences in order to devise possible solutions for complex political and moral problems: 'citizens from different walks of life have different experiences of problems and policies of public interest, experiences that have evidential import for devising and evaluating solutions... inclusion makes maximal use of such situated knowledge' (E. Anderson

2006, 14).<sup>123</sup> Exclusions may be justified, however, where an individual or group is committed to anti-social principles that threaten the very democratic ideal that motivates inclusivity:

Intolerance, abuse, calling of names because of differences of opinion about religion or politics or business, as well as because of differences of race, color, wealth or degree of culture are treason to the democratic way of life. For everything which bars freedom and fullness of communication sets up barriers that divide human beings into sets and cliques, into antagonistic sects and factions, and thereby undermines the democratic way of life. (Dewey 1988, 227-28)

The classical pragmatist understanding of social and moral inquiry, then, is an experimental procedure that ideally incorporates a diverse range of perspectives in determining the nature of different moral and political problems, the potential principles or policies that may address these problems, and then tests these hypothesized solutions in practice.

Anderson uses the pragmatist conception of inquiry as a framework for understanding how widespread moral beliefs change within society, and how to determine when those changes constitute moral progress. Using the abolition of slavery as a case study, Anderson shows that this important moral improvement came about not simply because of the many arguments advanced by abolitionist social theorists, philosophers, and theologians. Rather, the abolitionist movement was successful in producing moral progress because it followed the method of inquiry suggested by the pragmatists, particularly in using methods that Anderson describes as bias correction and experiments in living. Taken together, these strategies yield an account that we can use to understand the social processes by which our concepts change and through which we regain conceptual confidence. These activities thus

---

<sup>123</sup> For a similar argument for a diverse community of inquiry, drawing on feminist epistemology rather than pragmatism, see Anderson (1995a).

represent an important range of cases where grassroots conceptual engineering has been successful in unsettling widespread conceptual incapacities.

The first method, *bias correction*, operates in the initial stages of moral inquiry. When reflection on our moral judgements reduces our confidence in a moral principle, we employ this method to '[aid] in the formulation or revision of moral principles and norms prior to their implementation' (E. Anderson 2014a, 6). Anderson describes several ways bias correction can be achieved, through the example of the abolitionist movement's 'contentious politics'.<sup>124</sup> This term refers to a kind of political action whereby 'people make claims against others, on behalf of someone's interests' (E. Anderson 2015b, 32).<sup>125</sup> Many different practices are captured by this label, ranging from philosophical argument to peaceful protest to civil disobedience to rioting to civil war. After a group of people becomes convinced that some widespread practice is morally wrong, 'contention consists in the collective, concerted repudiation of morally objectionable practices by means of actions that disrupt the routine functioning of those practices, and that express rejection of the moral authority of people to practice them' (33). In doing so, participants in the contention prompt others to stop participating in what is now a controversial practice and to reflect critically on their moral principles, by holding people accountable for their participating in the disputed practice (E. Anderson 2014a, 10). That is to say, contentious politics aims to hold people accountable for their problematic moral judgements and principles, and demands that they take responsibility for them: 'Actual confrontation with a mass of claims that conflict with the reigning norm forces them to engage in *practical reflection*

---

<sup>124</sup> Anderson takes the term 'contentious politics' from social theory. See Tilly (2005), Tilly and Tarrow (2006).

<sup>125</sup> In addition to contentious politics, Anderson elsewhere argues for institutional methods of correcting bias. Where formal organizations are involved in the inquiry, such as governments or universities, they should take steps to ensure that people from diverse backgrounds all have an opportunity to contribute, by, for example, desegregating communities and instituting affirmative action programmes (E. Anderson 1995b, 2010, 2012).

as to whether they should continue to support the norm' (13). This serves to prompt those in positions of power to consider alternative perspectives by making it difficult for them to continue to impose their will (E. Anderson 2015b, 39). When a large and unified movement demonstrates its worthiness and commitment to a cause, it becomes difficult for the status quo to be maintained.

The specific form of contentious politics practised by the abolitionists was that of the social movement. British abolitionists became organized in 1787. They recruited like-minded activists, distributed a single logo for their group, maintained a mailing list to keep their members and supporters informed of key events, set up action committees, boycotted products, took out advertisements, published pamphlets and books, issued 'report cards' on their elected representatives' support for slavery, and organized the first mass-scale petition and litigation campaigns in British history (E. Anderson 2014a, 11). Their actions contributed to the change in public opinion that led to the criminalization of slaveholding and slave-trading in (most of) the British Empire in 1834.<sup>126</sup> But the actions of white abolitionists alone were not sufficient to correct the relevant biases. As black abolitionists pointed out, even well-intentioned white abolitionists often held some racial bias, which sometimes worked against the movement's own interests by undermining their arguments. In order to counteract this bias within and without the abolitionist movement, 'much more than pure moral argument was required. Blacks needed to *demonstrate in action* their *interest, capacity, and worthiness for freedom and dignity*' (Anderson 2015a, 36, emphasis hers). Black abolitionists, mostly former slaves themselves, did so through eloquent speeches and writings, showing that they had just as much human worth as others and at the same time decrying the brutalities of slavery and the racist biases that were prevalent even among white abolitionists.

---

<sup>126</sup> Slavery remained legal in territories administered by the East India Company until 1843.

The second method, *experiments in living*, constitutes the phase of moral inquiry in which new moral principles are tested. After the social process of bias correction has identified candidates for new or revised moral principles, ‘we can test our moral principles by putting them into practice and seeing whether we can live with the results... When we put them into practice, we can determine whether they solve the problem they were supposed to solve, with acceptable side effects’ (E. Anderson 2014b, 5).<sup>127</sup> In the case of slavery, there were two problems that abolitionists needed to demonstrate could be surmounted in the absence of slavery. The first was ‘the assumption that slaves were unfit for freedom’ which therefore required a test of ‘how the freed people exercised their freedom’ (E. Anderson 2015c, 87). The second was the problem of how to establish a labour regime that would produce sufficient surplus goods to support other social institutions without relying on forced labour, or what Anderson calls the ‘problem of civilization’ (E. Anderson 2014b). At the time, the abolitionist experiment was seen as a failure: cash crop yields plummeted as former slaves worked fewer hours and focused their labour more on farming crops for domestic consumption than producing cotton and sugar for export. This result led many white critics to conclude that the experiment had merely confirmed that Africans were unsuited for civilized life and that only forced labour could ensure an efficient agricultural workforce. But this interpretation of the results of the abolitionist experiment was still clouded by racial bias. White critics of the new labour regime failed to take into account the former slaves’ own justification for how they exercised their newfound freedom. In Haiti, for example, the former slaves diversified their agriculture to produce crops needed to support a population that had been underfed and malnourished under the slaver regime (E. Anderson 2015c, 90). Moreover, while

---

<sup>127</sup> See Anderson (1991) for a discussion of the Millian idea of experiments in living. Anderson argues that we can read Mill’s misery during his early life as a Benthamite as an experiment refuting Bentham’s hedonistic utilitarianism.

freed black farmers had indeed reduced their daily working hours, they did so only to the point that their workday was on a par with that of white labourers. Free workers were in fact *more* productive than slaves, but only during work hours; they simply worked fewer hours than they were forced to under slavery, which routinely worked slaves to death (E. Anderson 2014b). A proper assessment of the experiment in living represented by abolition would have revealed the test to be a success, but further bias correction was needed before this conclusion could become clear to many Europeans.

Anderson suggests that the best way we have to judge that we have actually made moral progress, without begging the question in favour of our present moral principles, is to ensure that the moral inquiries that led us to our present values have *removed* bias, partiality, ignorance, and inconsistency. For instance, proslavery arguments were 'based on a massive confusion on the part of the powerful between the morally right and their own self-interest', while 'the abolitionist social movement possessed features that... would tend to correct the ignorance and confusion of the powerful' (E. Anderson 2014a, 23). Because supporters of slavery had moral principles muddled by racial bias, self-interest, ignorance of black experiences, and confusion regarding free labour, we have reason to believe that a moral change that removed some or all of these moral-epistemological defects would constitute an improvement. The change brought about by abolitionism constitutes just such a shift, so we have reason to believe that the abolition of slavery indeed brought about moral progress.

Above, I suggested that the cycle through which we lose and regain confidence in our concepts can be modelled after the pragmatist understanding of inquiry. Anderson's understanding of moral inquiry is particularly useful for this purpose, because she not only describes some of the social practices by which a community can undermine and then return to ethical confidence, but also gives us a way to determine whether that newfound (or refounded) confidence is warranted. When a situation such as those described in the first section of this chapter arises often enough, that will



prompt the members of the community to critically reflect upon their concepts and lose their conceptual confidence. Forms of contentious politics and mechanisms of accountability serve to spread awareness of the identified conceptual incapacities as such, prompting more members of a wider community to reflect. At the same time, new or revised concepts and conceptions may be invented by looking for and correcting potential biases; this step must be done with care in order to ensure that a suitably diverse range of perspectives contributes to our revision of concepts and conceptions. Along with the concomitant changes in the community's social practices, these new concepts and conceptions are then put to the test in order to see whether the conceptual changes bring a return to confidence — or else the test is to see whether we can carry on after rejecting the old concepts without a replacement. Their emerging confidence in those new concepts and conceptions should then be investigated to ensure that it constitutes a genuine improvement — to see whether the new concepts and conceptions actually solve the ethical, epistemic, or social problems posed by the old concepts and conceptions; whether they raise new problems; and whether their institution removes bias, partiality, or inconsistency.

Conceptual inquiry can go on at a highly theoretical level, as when theorists and philosophers suggest new concepts for making distinctions that may be more useful to scientific experiment and observation (oxygen rather than phlogiston), for mathematical systems (classical understandings of infinity as a boundless limit vs. the Cantorian conception of transfinite numbers), or to philosophical theories (time as A-series or B-series). Indeed, many conceptual engineering projects go on in this theoretical domain, where expert judgement bears considerable weight. But the kinds of concepts and conceptions with which I and those engaged in ameliorative inquiry have been concerned, namely, social and moral concepts, are more complicated objects of conceptual inquiry. This is because, as we have seen, social and moral concepts are embedded in social practices with concrete effects on the lives of the people who use

them and are categorized by them. And because the same domain of concepts is involved in evaluating the results of our experiments with different concepts and conceptions, Williams correctly notes that it can be difficult to judge whether our new concepts and conceptions constitute an improvement. Anderson's model of moral inquiry is helpful, then, because it provides a way to understand how we can bring about social and moral improvements that may be assessed as such.

Hermeneutical dissent can be viewed as the paradigmatic case of conceptual inquiry in this domain, because it incorporates the methods of bias correction and experiments in living. Recall that, in Chapter 2, I defined hermeneutical dissent as a marginalized community's organized efforts to invent and disseminate concepts and conceptions that capture distinctive and important aspects of their social experience. Experts, such as conceptual engineers, certainly have a role to play in this process, but on the pragmatist model of conceptual inquiry, experts are neither the only inquirers nor the sole authorities. We can see this in the examples I discussed earlier. For instance, the concept *SEXUAL HARASSMENT* did not exist in the mid-twentieth century, just when changing social conditions — women's greater presence in the workforce and in public life — had made this social experience more common, or at least more noticeable. But because the concept of sexual harassment was absent in the collective hermeneutical resource, it was difficult for those who experienced it to express what happened to them without their experience being interpreted in a way biased towards men's perspectives, such as flirtation. Feminists such as Carmita Wood organized to develop an understanding of their experiences that corrected for the bias in favour of men's perspectives, identifying sexual harassment by name and calling attention to its wrongfulness and prevalence. Their activism around the issue — protests, lobbying, lawsuits, speeches, consciousness-raising groups, and publications — was a form of contentious politics that aimed both at holding sexual harassers accountable and at changing widespread interpretations of the issue by explaining the nature of sexual

harassment. The enforcement of corporate and legal policies regarding sexual harassment can be regarded as an experiment in living: now that we have taken up the concept of sexual harassment, we have had to change our social practices and institutions. Removing a bias in favour of men's perspectives by introducing a new concept has improved the lives of women by making it a default presumption, in many places, that sexual harassment will not be tolerated.

A range of similar experiments is underway in LGBTQ+ communities, as queer activists develop and debate a multitude of concepts that may be used to capture their social and personal identities. The concept AGENDER and related concepts such as NON-BINARY GENDER, for example, result from removing the presumption that all human beings have a male or female gender identity. If this pair of concepts cannot accommodate all gender identities, then we need one or more new concepts to fill the gap. Because the presumption in favour of a gender binary privileges heterosexuality as the norm, introducing non-binary identities aims to correct for a bias. The contentious politics of non-binary activists is illustrated by the work of Tyler Ford, who writes for magazines, gives interviews, and participates in protests calling attention to non-binary, trans, and other LGBTQ+ issues. Another way agender people call out heteronormative bias is by asking to be referred to by singular they/them pronouns or nonstandard personal pronouns such as 'xe' or 'zhe'. The ongoing experiment in living that makes room for non-binary genders involves institutional change — such as allowing identification documents to carry a gender label other than 'F' or 'M' — as well as changes to interpersonal interactions — such as becoming comfortable with seeing people whose appearance does not neatly fall into one gender category or another.

The pragmatist account of inquiry can also explain why removing certain concepts from use, such as SAVAGE, constitutes a moral and epistemic improvement. First, the very creation of categories that license discrimination on the basis of

membership in a stereotypically defined social group incorporates a bias in favour of the socially powerful, providing them with a ready justification for enforcing their will against those in the target group. Second, the stereotypical description involved in concepts such as *SAVAGE* typically involve a number of falsehoods, for instance, that indigenous people lack self-control or are simple-minded. Third, even the commitments to true propositions involved in this stereotype are problematic because of the noxious inferences they appear to license. For instance, the fact that indigenous people are overrepresented in the Canadian prison population is sometimes used in inferences to the conclusion that First Nations people are inherently prone to violence or criminality, a conclusion consonant with the negative evaluative aspect of *SAVAGE*. Fourth, we can look to history for the results of an experiment in living that show that when indigenous people are conceived as ‘savages’, the abuses and atrocities committed against them are legion. Altogether, rejecting the concept *SAVAGE* serves to correct bias, remove ignorance and error, and acknowledge the historical facts. This is more than enough reason to think that *SAVAGE* constitutes a conceptual incapacity, and its removal from use (other than to interpret the thoughts and actions of those who used it in the past or who continue to use it) constitutes moral and epistemic progress.

The pragmatist method also gives us reason to reject noxious new concepts and conceptions when they appear, reasserting our confidence in our old concepts. The concept of a ‘race traitor’, for example, was developed by white nationalists and white supremacists to identify and stigmatize those who have close (especially romantic or sexual) relations with someone from another racial group. But the production of such a concept does not correct for any bias or ignorance — just the opposite. White nationalists occupy a position of social privilege in virtue of their racial group (even if they are often disadvantaged along other axes, such as socioeconomic class). The concept of a ‘race traitor’ serves to maintain that social power by introducing barriers to the inclusion of racialized people in society. Moreover, the concept of a ‘race traitor’

depends upon a body of ignorance to function. To explain why it is a bad thing to fraternize with other racial groups, white nationalists typically appeal to outmoded ideas about traits that are supposedly inherent to different geographical or biological origins which will somehow be diluted if races are mixed. Finally, we can again use history as a kind of experiment in living, to see how similar ways of thinking about race relations have gotten on. And the record here is abysmal, incorporating colonialism, fascism, and Nazism. So we have plenty of reason to reject the concept RACE TRAITOR before it even has time to inspire confidence.

Institutional solutions will also play a part in effecting conceptual change. The existence of hermeneutical injustice means that the kind of communication between different social groups that is required to meet the diversity requirement of pragmatist inquiry does not obtain. As we have seen, diversity is needed for bias correction. Institutions can play a role in breaking down barriers to communication by changing policies to enable more diversity within themselves, or by using their influence to facilitate communication between different groups elsewhere. Newspapers can deliberately give column space to marginalized writers, who can explain to a wide audience their social experiences and the need for new concepts and conceptions to interpret them. For instance, *The Guardian* has done so in publishing Ford's autobiographical piece on their struggles with gender (Ford 2015). Universities can engage in deliberate efforts to diversify their faculty, student body, and participants in academic events such as conferences. The guidelines developed by the British Philosophical Association and the Society for Women in Philosophy, UK have informed numerous academic events and university departments in the UK to this end (The British Philosophical Association n.d.). Governments can also take action to desegregate educational institutions to expose all students to different viewpoints they might not otherwise encounter. Segregation is *de facto* a persistent issue in US schools,

despite having been *de jure* abolished in the 1950s, and contributes to the persistence of biases that contribute to hermeneutical injustices (E. Anderson 2012).

Taking responsibility for our concepts involves taking part in conceptual inquiry. When one encounters one of the scenarios described in §5.1, this may be a sign that one has a conceptual incapacity, which, as I argued in Chapter 4, places one under an obligation to take responsibility for one's concepts through critical reflection. We are now in a position to see more of what this reflection entails. It is not just an individual matter of scrutinizing one's concepts by thinking about one's own thinking. It is also a social matter, for conceptual inquiry requires taking diverse perspectives into account to correct for bias. In correcting for any applicable biases, we may deliberate on which concepts and conceptions we should be using. This may involve deference to the experts, who may be in a better position to do the conceptual engineering work. But this deliberation can also come from grassroots organizing aimed at generating acceptance or rejection of some concept or conception. Both theory and activism contribute to the contentious politics needed to shake the conceptual confidence of a widening circle of people by holding them accountable for continuing to use their received concepts. As more people take up the proposed conceptual change, they engage in experiments in living as social practices and institutions change in order to accommodate the new concepts and conceptions. Taking responsibility for one's concepts need not require taking an active role in either conceptual engineering or political activism. But it minimally requires us to make an effort to understand the conceptual changes being proposed, to consider their apparent merits, and to be prepared to offer a justification for not adopting the new concepts or conceptions should we choose to do so. And, should evidence mount that one's justification for sticking with the old concepts and conceptions is spurious, one should be prepared to change one's mind.

#### 5.4. Summary

In this chapter, I have explored the notion of taking responsibility for our concepts in more detail. I outlined various ways in which we may be prompted to reflect upon our concepts, which can precipitate a loss of confidence in them. I then suggested that one way to take responsibility for one's concepts is to defer to philosophers and expert theorists who are in a better position to engage in conceptual engineering. But because of the limitations of philosophical reflection, simply appealing to the experts is insufficient. I then turned to a pragmatist account of social and moral inquiry to explain how communities criticize and return to confidence in their concepts. Taking responsibility for our concepts is not just an individual act of reflection, but a social process in which we are required to take part.





## Conclusion

In this thesis, I have presented the basis for an account of the ethics of our conceptual practices. I have shown how the use of problematic concepts, and ignorance of or non-competence with concepts developed through hermeneutical dissent, can be morally, epistemically, and socially problematic. I have presented a theory of moral responsibility that shows that these conceptual incapacities can be attributed to us, such that we are responsible for them and open to appraisal for them. I have argued that we can dissolve problems for moral responsibility generated by bad moral luck and similar phenomena by distinguishing between the attributability and answerability senses of responsibility. The latter concerns requirements to take responsibility for various things that may not be attributable to us. With regard to our concepts, we are required to engage in critical reflection, in conversation with various lay and expert communities.

To return to the example with which I opened, the account I have presented enables us definitively to say that the Nazi propagandists were responsible for what they did. In twisting concepts like HEROISM and introducing other problematic concepts before and during their rule, the architects and distributors of the LTI may have acted in ignorance that they were doing wrong. They may have been firmly convinced that Hitler's view of the world was right. But they had ample evidence that should have led them to conclude otherwise. Perhaps some of them acted from some mental disorder or other, but it is massively implausible that the entire Nazi regime and the LTI can be explained in this way. One aspect of what is so terrifying about shifts to totalitarianism is that many of the regime's adherents and central participants seem to be acting perfectly rationally — just from some fundamentally misguided premises in which their confidence is unjustifiably strong. The LTI represents an evil misuse of some of the methods of conceptual engineering and conceptual inquiry to impose a deeply mistaken worldview and enact a regime not just of political, but cognitive control.

Those who would revive these ways of thinking today are similarly culpable, and deserve our ire.

The young people partially raised under Nazism that Klemperer sought to re-educate after the war are perhaps a more complicated story. As victims of a nationwide attempt at brainwashing, it is possible that their use of the concepts and conceptions of the LTI could, for a time, be excused. A person who has been so thoroughly brainwashed as to be unable, without deprogramming, to respond to the right reasons, is not responsible for what she does. Additionally, many of those who spoke and behaved in ways aligning with the regime's prescriptions did so under coercion: the price of dissent was imprisonment, torture, and death. If, in adopting the language of the LTI in public to protect themselves, these people were coerced into conceptual incapacities, they could be excused from fault. However, those who refused or simply failed to adjust their ways of thinking in the post-war period retained conceptual incapacities in the face of mounting evidence, and can thereby be found responsible. Klemperer's efforts to de-Nazify their thinking by explaining the poison of the LTI is one example of how educators have an important role to play in conceptual inquiry.

By way of closing, I wish to speculate briefly on a few future ways in which the notion of conceptual responsibility may be further developed, and its broader significance for academic philosophy.

To effect conceptual change for the better, and to guard against efforts to force conceptual change for the worse, we might wish to incorporate other methods of debiasing in our conceptual inquiries in addition to the methods of contentious politics discussed by Anderson. It is highly plausible that aligning an account of conceptual responsibility with a particular theory of concepts could help develop such methods. Psychological research could be employed, for example, to show how certain ways of debiasing are more or less effective in producing changes to the prototypes, exemplars,

and mental theories we use (if, indeed, that is what concepts are best understood to be). Understanding how conceptual change can be done psychologically, coupled with philosophical proposals for which concepts we should or should not encourage, may prove useful to efforts to improve our concepts. Such work might also reveal effective ways to resist the imposition of forms of cognitive control, such as the LTI.

A striking feature of the Nazi worldview, as observed by Hannah Arendt, is how it abused certain *scientific* concepts. Arendt describes the Nazi form of totalitarianism as founded on an ideology that aimed to explain all human activity by reference to *nature*, and adopted a misconception of natural selection as the basis for governance:

Underlying the Nazis' belief in race laws as the expression of the law of nature in man, is Darwin's idea of man as the product of a natural development which does not necessarily stop with the present species of human beings... If it is the law of nature to eliminate everything that is harmful or unfit to live, it would mean the end of nature itself if new categories of the harmful and unfit-to-live could not be found... In the body politic of totalitarian government, [the] place of positive laws is taken by total terror, which is designed to translate into reality the law of movement of... nature... Practically speaking, this means that terror executes on the spot the death sentences that Nature is supposed to have pronounced on races or individuals who are 'unfit to live' ... without waiting for the slower and less efficient processes of nature... themselves. (Arendt 1976, 463, 464, 466)

Of course, the role of scientific concepts in our lives is not always so dramatic. Scientific research has taken on an important role in our individual and social lives, usually construed in terms of the useful *knowledge* that science can provide.<sup>128</sup> But *concepts* inherited from the scientific context also play an increasingly important role in how we live. For example, Kristin Shrader-Frechette discusses a toxicological concept, *HORMESIS*, which refers to a purported phenomenon where low-level exposure to toxins has a positive health outcome (Shrader-Frechette 2010). She criticizes this concept as confused and often lacking basis in fact, and, referring to case studies, shows how the use of this concept has produced harmful regulations and legal decisions related to

---

<sup>128</sup> Cf. Kitcher (2001), Grasswick (2018).

workplace and domestic exposure to toxins. We might expect other scientific conceptual incapacities to be similarly problematic. Conceptual inquiry in the scientific domain is needed to help identify and rectify such problems.

In addition to relatively concrete social, ethical, and scientific concepts, it may be that we are required to take responsibility for some fairly abstract and general concepts. Throughout the thesis, I referred to issues concerning the First Nations of Canada, that is to say, the peoples who are indigenous to that part of North America. I did so most vividly with regard to the concept of a 'savage' in order to make the point about the wrongfulness of certain concepts. But this point ought to be obvious to anyone who disavows racism. As Canadian society and other colonial powers move to take responsibility for their abhorrent treatment of indigenous peoples, a potentially more radical shift in our concepts may take shape. In particular, if we are to reconcile indigenous and settler laws as part of honouring the treaties signed over the centuries, the concepts of LAW and JUSTICE themselves may require revision. Indigenous conceptions of justice as involving the reconciliation of wrongdoer and wronged, rather than punishment of the wrongdoer by the state, have been widely discussed. This difference between European and indigenous legal systems is profound. As a report commissioned by the Government of Manitoba argues, at the time of colonization, 'Europeans and Aboriginal people viewed the same crime of murder in different ways. The two groups perceived the other's system of justice as inconsistent, incoherent, and incomprehensible' (Aboriginal Justice Implementation Commission 1999, chap. 2). Moral and political reconciliation with indigenous peoples may require some *conceptual reconciliation* as well.<sup>129</sup> This process will involve more than token

---

<sup>129</sup> I was first struck by this possibility at a session of the 2018 Congress of the Canadian Philosophical Association on 'Truth and Reconciliation in Academia'. I thank Mike Giudice for further discussion of the idea of conceptual reconciliation.

efforts by academics and colonial and indigenous governments — conceptual reconciliation requires conceptual inquiry.

These possibilities for future work bring us back to the conception of philosophy I mentioned in the introduction. We have seen that philosophical reflection is but one part of conceptual inquiry, and sometimes a fairly minor one. Still, the possibility that philosophers are in a position to present well-grounded proposals for conceptual revision as part of conceptual inquiry is a significantly different view to that expressed by the adage that philosophy is concerned with the analysis and clarification of concepts, otherwise leaving everything as it is. Conceptual inquiry requires ongoing, socially engaged efforts to improve the ways in which we think. This requires not just philosophical argument, but also an understanding of how conceptual change can be effected, and attention to those areas where conceptual change is either ongoing or sorely needed. Shrader-Frechette describes philosophers of science as having ‘professional duties, like those of scientists and engineers... to protect the public’, as well as ‘justice-based duties to help correct the flawed science that is often used to justify... unfair risk distributions’ (Shrader-Frechette 2010, 464). Critiquing scientific concepts in research articles is only one part of this pair of duties; Shrader-Frechette describes how, in her own work, she writes popular pieces, engages in public service work with governments and NGOs, and involves students in the analysis of science-based policy. These activities, where they concern the pernicious use of scientific concepts, represent another way that philosophers and lay communities can collaborate on conceptual inquiry. In the light of the picture of conceptual inquiry I have presented, conceptual engineering requires a significant reimagining of what doing philosophy looks like.

Perhaps, as Philip Kitcher writes, we must turn the organization of professional philosophy inside-out, and focus the core of the discipline not on abstract and specialized conceptual puzzles, but instead on pressing conceptual questions of social

concern (Kitcher 2011). Or perhaps it would suffice for there simply to be more philosophers taking an active part in this sort of work. At any rate, if we are required to take responsibility for our concepts, and if philosophers have anything distinctive and valuable to contribute, we have reason to broaden our conception of what philosophy is and should be.

## Bibliography

- @Alyssa\_Milano. 2017. "If You've Been Sexually Harassed or Assaulted Write 'Me Too' as a Reply to This Tweet." Twitter. October 15, 2017.  
[https://twitter.com/alyssa\\_milano/status/919659438700670976](https://twitter.com/alyssa_milano/status/919659438700670976).
- @realdonaldtrump. 2012. "The Concept of Global Warming Was Created by and for the Chinese in Order to Make U.S. Manufacturing Non-Competitive." Twitter. November 6, 2012.  
<https://twitter.com/realdonaldtrump/status/265895292191248385>.
- Aboriginal Justice Implementation Commission. 1999. "Report of the Aboriginal Justice Inquiry of Manitoba — Volume I: The Justice System and Aboriginal People."  
<http://www.ajic.mb.ca/volumel/toc.html>.
- Adams, Robert Merrihew. 1985. "Involuntary Sins." *Philosophical Review* 94 (1): 3–31.
- American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. Arlington, Va: American Psychiatric Publishing.
- Anderson, Derek Egan. 2017. "Conceptual Competence Injustice." *Social Epistemology* 31 (2): 210–23. <https://doi.org/10.1080/02691728.2016.1241320>.
- Anderson, Elizabeth. 1991. "John Stuart Mill and Experiments in Living." *Ethics* 102 (1): 4–26. <https://doi.org/10.1086/293367>.
- . 1995a. "Feminist Epistemology: An Interpretation and a Defense." *Hypatia* 10 (3): 50–84. <https://doi.org/10.2307/3810237>.
- . 1995b. "The Democratic University: The Role of Justice in the Production of Knowledge." *Social Philosophy and Policy* 12 (2): 186–219.
- . 2006. "The Epistemology of Democracy." *Episteme* 3 (1–2): 8–22.
- . 2010. *The Imperative of Integration*. Princeton, NJ: Princeton University Press.
- . 2012. "Epistemic Justice as a Virtue of Social Institutions." *Social Epistemology* 26 (2): 163–73. <https://doi.org/10.1080/02691728.2011.652211>.
- . 2014a. "Social Movements, Experiments in Living, and Moral Progress: Case Studies from Britain's Abolition of Slavery." In *The Lindley Lecture*. Lawrence, KS: University of Kansas.
- . 2014b. "The Quest for Free Labor: Pragmatism and Experiments in Emancipation." In *The Amherst Lecture in Philosophy*, Lecture 9. Amherst, MA: Amherst College. <http://www.amherstlecture.org/anderson2014/>.
- . 2015a. "Feminist Epistemology and Philosophy of Science." Edited by Edward N Zalta. *The Stanford Encyclopedia of Philosophy*.  
<http://plato.stanford.edu/archives/fall2015/entries/feminism-epistemology/>.
- . 2015b. "Moral Bias and Corrective Practices: A Pragmatist Perspective." *Proceedings and Addresses of the American Philosophical Association* 89: 21–47.
- . 2015c. "The Social Epistemology of Morality: Learning from the Forgotten

- History of the Abolition of Slavery." In *The Epistemic Life of Groups: Essays in the Epistemology of Collectives*, edited by Michael S. Brady and Miranda Fricker, 75–94. Oxford: Oxford University Press.
- Anderson, Mark Cronlund, and Carmen L. Robertson. 2011. *Seeing Red: A History of Natives in Canadian Newspapers*. Winnipeg, MB: University of Manitoba Press.
- Annas, Julia. 1993. *The Morality of Happiness*. Oxford: Oxford University Press.
- Anscombe, G. E. M. 1958a. "Modern Moral Philosophy." *Philosophy* 33 (124): 1–19.
- . 1958b. "On Brute Facts." *Analysis* 18 (3): 69–72.
- Arendt, Hannah. 1976. *The Origins of Totalitarianism*. Orlando, FL: Houghton Mifflin Harcourt.
- Aristotle. 1925. *Nicomachean Ethics*. Translated by W. D. Ross. The Internet Classics Archive. <http://classics.mit.edu/Aristotle/nicomachaen.html>.
- Arpaly, Nomy. 2002. *Unprincipled Virtue*. Oxford, UK: Oxford University Press.
- . 2006. *Merit, Meaning, and Human Bondage*. Princeton, NJ: Princeton University Press.
- Baier, Kurt. 1987. "Moral and Legal Responsibility." In *Medical Innovation and Bad Outcomes: Legal, Social and Ethical Responses*, edited by Mark Siegler, Stephen Toulmin, Franklin E. Zimring, and Kenneth F. Schaffner, 101–29. Ann Arbor, MI: Health Administration Press.
- Begby, Endre. 2013. "The Epistemology of Prejudice." *Thought* 2: 90–99. <https://doi.org/10.1002/tht3.71>.
- Berlin, Isaiah. 1999. "The Purpose of Philosophy." In *Concepts and Categories: Philosophical Essays*, edited by Henry Hardy, 1–11. London: Pimlico.
- Blackburn, Simon. 1998. *Ruling Passions*. Oxford, UK: Clarendon Press.
- . 2013. "Disentangling Disentangling." In *Thick Concepts*, edited by Simon Kirchin, 121–35. Oxford: Oxford University Press.
- Blum, Lawrence. 2007. "Racial Virtues." In *Working Virtue: Virtue Ethics and Contemporary Moral Problems*, edited by Rebecca L. Walker and Phiip J. Ivanhoe, 225–49. Oxford, UK: Clarendon Press. [http://faculty.umb.edu/lawrence\\_blum/publications/publications/A51.pdf](http://faculty.umb.edu/lawrence_blum/publications/publications/A51.pdf).
- Bohman, James. 2010. "Ethics as Moral Inquiry: Dewey and the Moral Psychology of Social Reform." In *The Cambridge Companion to Dewey*, edited by Molly Cochran, 187–210. Cambridge, UK: Cambridge University Press.
- Brandom, Robert B. 1998. *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, MA: Harvard University Press.
- . 2009. *Articulating Reasons: An Introduction to Inferentialism*. Cambridge, MA: Harvard University Press.
- Brownmiller, Susan. 1990. *In Our Time: Memoir of a Revolution*. New York: Dial Press.



- Brownstein, Michael, and Jennifer Saul, eds. 2016a. *Implicit Bias and Philosophy, Volume 1: Metaphysics and Epistemology*. Oxford: Oxford University Press.
- . , eds. 2016b. *Implicit Bias and Philosophy, Volume 2: Moral Responsibility, Structural Injustice, and Ethics*. Oxford: Oxford University Press.
- Burgess, Alexis, and David Plunkett. 2013a. "Conceptual Ethics I." *Philosophy Compass* 8 (12): 1091–1101.
- . 2013b. "Conceptual Ethics II." *Philosophy Compass* 8 (12): 1102–10.
- Burke, Tarana. n.d. "Me Too." Metoomvmt.Org. Accessed June 13, 2018. <http://metoomvmt.org/>.
- Cappelen, Herman. 2018. *Fixing Language: An Essay on Conceptual Engineering*. Oxford: Oxford University Press.
- Carey, Susan. 1985. *Conceptual Change in Childhood*. Cambridge, MA: MIT Press.
- Carruthers, Peter. 1987. "Conceptual Pragmatism." *Synthese* 73 (2): 205–24. <https://doi.org/10.1007/BF00484740>.
- Cassam, Quassim. Forthcoming. "Epistemic Insouciance." *Journal of Philosophical Research*.
- Chappell, Sophie-Grace. 2013. "There Are No Thin Concepts." In *Thick Concepts*, edited by Simon Kirchin, 182–95. Oxford: Oxford University Press.
- Collins, Patricia Hill. 2009. *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. 2nd ed. New York: Routledge.
- Congdon, Matthew. 2015. "Epistemic Injustice in the Space of Reasons." *Episteme* 12 (1): 75–93. <https://doi.org/10.1017/epi.2014.34>.
- Cornford, Francis MacDonald. 1912. *From Religion to Philosophy: A Study in the Origins of Western Speculation*. New York: Longmans Green.
- Craig, Edward. 1999. *Knowledge and the State of Nature: An Essay in Conceptual Synthesis*. Oxford: Oxford University Press. <https://doi.org/10.1093/0198238797.001.0001>.
- Crerar, Charlie. 2016. "Taboo, Hermeneutical Injustice, and Expressively Free Environments." *Episteme* 13 (2): 195–207. <https://doi.org/10.1017/epi.2015.35>.
- Darwall, Stephen. 2006. *The Second-Person Standpoint: Morality, Respect, and Accountability*. Cambridge, MA: Harvard University Press.
- Deleuze, Gilles, and Félix Guattari. 1994. *What Is Philosophy?* Translated by Hugh Tomlinson and Graham Burchill. New York: Columbia University Press.
- Dembroff, Robin A. 2016. "What Is Sexual Orientation?" *Philosophers' Imprint* 16 (3): 1–27.
- Dewey, John. 1980. *Democracy and Education (The Middle Works of John Dewey, 1899–1924. Vol. 9)*. Edited by Jo Ann Boydston, Patricia R. Baysinger, and Barbara Levine. Carbondale, IL: Southern Illinois University Press.

- . 1983. *Human Nature and Conduct (The Middle Works of John Dewey, 1899–1924. Vol. 14)*. Edited by Jo Ann Boydston and Patricia Baysinger. Carbondale, IL: Southern Illinois University Press.
- . 1984. “The Public and Its Problems.” In *The Later Works of John Dewey, 1925–1953. Vol. 2: 1925–1927, Essays, The Public and Its Problems*, edited by Jo Ann Boydston and Bridget A. Walsh, 235–372. Carbondale, IL: Southern Illinois University Press.
- . 1986. *Logic: The Theory of Inquiry (The Later Works of John Dewey, 1925–1953. Vol. 12)*. Edited by Jo Ann Boydston and Kathleen Poulos. Carbondale, IL: Southern Illinois University Press.
- . 1988. “Creative Democracy—The Task Before Us.” In *The Later Works of John Dewey, 1925–1953. Vol. 14: 1939–1941, Essays*, edited by Jo Ann Boydston and Anne Sharpe, 224–30. Carbondale, IL: Southern Illinois University Press.
- Dewey, John, and James Hayden Tufts. 1985. *Ethics, Revised Edition (The Later Works of John Dewey, 1925–1953. Vol. 7)*. Edited by Jo Ann Boydston and Barbara Levine. Carbondale, IL: Southern Illinois University Press.
- Dotson, Kristie. 2012. “A Cautionary Tale: On Limiting Epistemic Oppression.” *Frontiers: A Journal of Women’s Studies* 33 (1): 24–47.
- Doucet, Mathieu, and Dylon McChesney. ms. “Culpable Ignorance and Mental Disorders.” Unpublished Manuscript.
- Dummett, Michael. 1993. *Frege: Philosophy of Language*. 2nd ed. Cambridge, MA: Harvard University Press.
- . 1996. *Seas of Language*. Oxford: Oxford University Press.
- . 2010. *The Nature and Future of Philosophy*. New York: Columbia University Press.
- Enoch, David. 2012. “Being Responsible, Taking Responsibility, and Penumbral Agency.” In *Luck, Value, and Commitment: Themes From the Ethics of Bernard Williams*, edited by Ulrike Heuer and Gerald Lang, 95–131. Oxford: Oxford University Press.
- Enoch, David, and Andrei Marmor. 2007. “The Case against Moral Luck.” *Law and Philosophy* 26 (4): 405–36. <https://doi.org/10.1007/s10982-006-9001-3>.
- Eshleman, Andrew. 2016. “Moral Responsibility.” Edited by Edward N Zalta. *The Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/archives/win2016/entries/moral-responsibility/>.
- Fanon, Frantz. 2008. “The Fact of Blackness.” In *Black Skin, White Masks*, translated by Charles Lam Markmann, 82–108. London: Pluto Press.
- Feibleman, James. 1943. “A Systematic Presentation of Peirce’s Ethics.” *Ethics* 53 (2): 98–109.
- Feinberg, Joel. 1970. *Doing and Deserving: Essays in the Theory of Responsibility*. Princeton, NJ: Princeton University Press.

- Filipovic, Jill. 2018. "The Poorly Reported Aziz Ansari Exposé Was a Missed Opportunity." *The Guardian*, January 16, 2018.  
<https://www.theguardian.com/commentisfree/2018/jan/16/aziz-ansari-story-missed-opportunity>.
- Fischer, John Martin, and Mark Ravizza. 1998. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge, UK: Cambridge University Press.
- FitzPatrick, William J. 2008. "Moral Responsibility and Normative Ignorance: Answering a New Skeptical Challenge." *Ethics* 118 (4): 589–613.
- Fodor, Jerry. 1975. *The Language of Thought*. Cambridge, MA: Harvard University Press.
- . 1981. "The Present Status of the Innateness Controversy." In *Representations: Philosophical Essays on the Foundations of Cognitive Science*. Cambridge, MA: MIT Press.
- . 1998. *Concepts: Where Cognitive Science Went Wrong*. Oxford: Oxford University Press.
- Foot, Philippa. 1954. "When Is a Principle a Moral Principle?" *Proceedings of the Aristotelian Society, Supplementary Volumes* 28 (Belief and Will): 95–110.
- . 2001. *Natural Goodness*. Oxford, UK: Oxford University Press.
- Ford, Tyler. 2015. "My Life without Gender." *The Guardian*, August 7, 2015.  
<https://www.theguardian.com/world/2015/aug/07/my-life-without-gender-strangers-are-desperate-to-know-what-genitalia-i-have>.
- Foucault, Michel. 1988. *Madness and Civilization: A History of Insanity in the Age of Reason*. Translated by Richard Howard. New York: Vintage.
- . 1990. *The History of Sexuality, Volume 1: An Introduction*. Translated by Robert Hurley. New York: Vintage.
- . 1995. *Discipline and Punish: The Birth of the Prison*. Translated by Alan Sheridan. New York: Vintage.
- . 2003. "The Subject and Power." In *The Essential Foucault: Selections from "Essential Works of Foucault: 1954–1984,"* edited by Paul Rabinow and Nikolas Rose, 126–44. New York: The New Press.
- Frankfurt, Harry G. 1971. "Freedom of the Will and the Concept of a Person." *The Journal of Philosophy* 68 (1): 5–20.
- . 2002. "Reply to Michael E. Bratman." In *The Contours of Agency: Essays on Themes from Harry Frankfurt*, edited by Sarah Buss and Lee Overton, 86–90. Cambridge, MA: MIT Press.
- . 2005. *On Bullshit*. Princeton, NJ: Princeton University Press.
- Freeman, Hadley. 2018. "What's the Difference between a Bad Date and a Wild Night? It Depends Who You Ask." *The Guardian*, January 27, 2018.  
<https://www.theguardian.com/culture/2018/jan/27/difference-bad-date-wild-night-aziz-ansari>.

- Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.
- . 2016. "Epistemic Injustice and the Preservation of Ignorance." In *The Epistemic Dimensions of Ignorance*, edited by Rik Peels and Martijn Blaauw, 160–77. Cambridge, UK: Cambridge University Press. <https://doi.org/10.1017/9780511820076>.
- . 2018. "Ambivalence About Forgiveness." In *Harms and Wrongs in Epistemic Practice*, edited by Simon Barker, Charlie Crerar, and Trystan S. Goetze. Cambridge, UK: Cambridge University Press.
- Friedersdorf, Conor. 2018. "How #MeToo Can Probe Gray Areas With Less Backlash." *The Atlantic*, January 18, 2018. <https://www.theatlantic.com/entertainment/archive/2018/01/how-metoo-can-probe-gray-areas-with-less-backlash/550559/>.
- Frye, Marilyn. 1983. "On Being White: Towards a Feminist Understanding of Race and Race Supremacy." In *The Politics of Reality: Essays in Feminist Theory*, 110–27. Berkeley, CA: Crossing Press.
- Garcia, Ernesto V. 2002. "A Kantian Theory of Evil." *The Monist* 85 (2): 194–209.
- Glover, Jonathan. 1970. *Responsibility*. New York: Humanities Press.
- Goetze, Trystan S. 2018. "Hermeneutical Dissent and the Species of Hermeneutical Injustice." *Hypatia* 33 (1): 73–90. <https://doi.org/10.1111/hypa.12384>.
- Goldberg, Michelle. 2014. "What Is a Woman? The Dispute between Radical Feminism and Transgenderism." *The New Yorker*, August 4, 2014. <https://www.newyorker.com/magazine/2014/08/04/woman-2>.
- Goldman, Alvin I. 2001. "Experts: Which Ones Should You Trust?" *Philosophy and Phenomenological Research* 63 (1): 85–110. <https://doi.org/10.1111/j.1933-1592.2001.tb00093.x>.
- Grasswick, Heidi. 2018. "Understanding Epistemic Trust Injustices and Their Harms." In *Harms and Wrongs in Epistemic Practice*, edited by Simon Barker, Charlie Crerar, and Trystan S. Goetze. Cambridge, UK: Cambridge University Press.
- Gzowski, Peter. 1978. "90 Minutes Live." Canada: Canadian Broadcasting Corporation. <http://www.cbc.ca/archives/entry/margaret-atwood-closet-cartoonist>.
- Hamblin, James. 2018. "This Is Not a Sex Panic." *The Atlantic*, January 17, 2018. <https://www.theatlantic.com/entertainment/archive/2018/01/this-is-not-a-sex-panic/550547/>.
- Hames, J. L., J. D. Ribeiro, A. R. Smith, and T. E. Joiner. 2012. "An Urge to Jump Affirms the Urge to Live: An Empirical Examination of the High Place Phenomenon." *Journal of Affective Disorders* 136 (2): 1114–20. <https://doi.org/10.1016/j.jad.2011.10.035>.
- Hare, R. M. 1988. "Comments." In *Hare and Critics*, edited by Douglas Seanor and N. Fotion, 199–294. Oxford: Clarendon Press.
- Harthan, Todd. 2015. "Rosewood." USA: 20th Television.

- Haslanger, Sally. 2012a. "Gender and Race: (What) Are They? (What) Do We Want Them To Be?" In *Resisting Reality: Social Construction and Social Critique*, 221–47. Oxford: Oxford University Press.
- . 2012b. *Resisting Reality: Social Construction and Social Critique*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199892631.001.0001>.
- . 2012c. "What Are We Talking About? The Semantics and Politics of Social Kinds." In *Resisting Reality: Social Construction and Social Critique*, 365–80. Oxford: Oxford University Press.
- . 2012d. "What Good Are Our Intuitions? Philosophical Analysis and Social Kinds." In *Resisting Reality: Social Construction and Social Critique*, 381–405. Oxford: Oxford University Press.
- Hieronymi, Pamela. 2014. "Reflection and Responsibility." *Philosophy and Public Affairs* 42 (1): 3–41.
- Hill, Anita F. 1997. *Speaking Truth to Power*. New York: Anchor Books.
- Hom, Christopher. 2008. "The Semantics of Racial Epithets." *Journal of Philosophy* 105 (8): 416–40.
- . 2012. "A Puzzle about Pejoratives." *Philosophical Studies* 159 (3): 383–405. <https://doi.org/10.1007/s11098-011-9749-7>.
- Hume, David. 1978. *A Treatise of Human Nature*. Oxford, UK: Oxford University Press.
- Ignatiev, Noel, and John Garvey, eds. 1996. *Race Traitor*. London: Routledge.
- Jackendoff, Ray. 1983. *Semantics and Cognition*. Cambridge, MA: MIT Press.
- Jenkins, Katharine. 2016. "Amelioration and Inclusion: Gender Identity and the Concept of Woman." *Ethics* 126 (2): 394–421. <https://doi.org/10.1086/683535>.
- Johnson, Casey Rebecca. 2018a. "For the Sake of Argument: The Nature and Extent of Our Obligation to Voice Disagreement." In *Voicing Dissent: The Ethics and Epistemology of Making Disagreement Public*, edited by Casey Rebecca Johnson. New York: Routledge.
- . 2018b. "Just Say 'No': Obligations to Voice Disagreement." In *Harms and Wrongs in Epistemic Practice*, edited by Simon Barker, Charlie Crerar, and Trystan S. Goetze. Cambridge University Press.
- Kenny, Anthony. 2010. "Concepts, Brains, and Behaviour." *Grazer Philosophische Studien* 81 (1): 105–13.
- Kirchin, Simon. 2013. "Introduction: Thick and Thin Concepts." In *Thick Concepts*, edited by Simon Kirchin, 1–19. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199672349.001.0001>.
- Kitcher, Philip. 2001. *Science, Truth, and Democracy*. Oxford, UK: Oxford University Press.
- . 2011. "Philosophy Inside Out." *Metaphilosophy* 42 (3): 248–60. <https://doi.org/10.1111/j.1467-9973.2011.01684.x>.

- Klemperer, Victor. 2002. *The Language of the Third Reich: LTI — Lingua Tertii Imperii: A Philologist's Notebook*. Translated by Martin Brady. London: Continuum.
- Lamb, Kate. 2018. "Stripped Naked and Beaten: Transgender Women Flee Indonesian City in Fear." *The Guardian*, February 2, 2018.  
<https://www.theguardian.com/world/2018/feb/02/gender-re-education-transwomen-flee-indonesia-aceh>.
- Langton, Rae. 2018. "Blocking as Counter-Speech." In *New Work on Speech Acts*, edited by Daniel Fogal, Daniel W Harris, and Matt Moss. Oxford: Oxford University Press.
- Laurence, Stephen, and Eric Margolis. 1997. "Regress Arguments Against the Language of Thought." *Analysis* 57 (1): 60–66.
- . 1999. "Concepts and Cognitive Science." In *Concepts: Core Readings*, edited by Eric Margolis and Stephen Laurence, 3–81. Cambridge, MA: MIT Press.
- Levy, Neil. 2014. *Consciousness and Moral Responsibility*. Oxford: Oxford University Press.
- Liptak, Adam. 2015. "Gender Bias Issue Could Tip Chief Justice Roberts Into Ruling for Gay Marriage." *The New York Times*, April 29, 2015. <http://nyti.ms/1OD08RL>.
- Locke, John. 1975. *An Essay Concerning Human Understanding*. Oxford, UK: Oxford University Press.
- Lopez, German. 2017. "Transgender People: 10 Common Myths." Vox. 2017.  
<https://www.vox.com/cards/transgender-myths-fiction-facts/transgender-people-confused-fiction>.
- Lorde, Audre. 2007. "Age, Race, Class, and Sex: Women Redefining Difference." In *Sister Outsider: Essays and Speeches*, 114–23. New York: Ten Speed Press.
- Lovibond, Sabina. 2015. "Ethical Upbringing: From Connivance to Cognition." In *Essays on Ethics and Feminism*, 128–45. Oxford: Oxford University Press.  
<https://doi.org/10.1093/acprof:oso/9780198719625.001.0001>.
- Lyons, Kate. 2016. "Transgender Stories: 'People Think We Wake up and Decide to Be Trans.'" *The Guardian*, July 10, 2016.  
<https://www.theguardian.com/society/2016/jul/10/transgender-stories-people-think-we-wake-up-and-decide-to-be-trans>.
- MacDonald, Michael. 2018. "3 Adults in Polyamorous Relationship Declared Legal Parents by N. L. Court." *CBC News*, June 14, 2018.  
<https://www.cbc.ca/news/canada/newfoundland-labrador/polyamorous-relationship-three-parents-1.4706560>.
- Machery, Edouard. 2009. *Doing Without Concepts*. Oxford: Oxford University Press.
- MacIntyre, Alasdair. 1998. *A Short History of Ethics: A History of Moral Philosophy from the Homeric Age to the Twentieth Century*. 2nd ed. London: Routledge.
- Margolis, Eric, and Stephen Laurence. 2007. "The Ontology of Concepts—Abstract Objects or Mental Representations?" *Noûs* 41 (4): 561–93.  
<https://doi.org/10.1111/j.1468-0068.2007.00663.x>.

- . 2014. "Concepts." Edited by Edward N Zalta. *Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/archives/spr2014/entries/concepts/>.
- Martino, Wayne. 1999. "'Cool Boys', 'Party Animals', 'Squids' and 'Poofters': Interrogating the Dynamics and Politics of Adolescent Masculinities in School." *British Journal of Sociology of Education* 20 (2): 239–63. <https://doi.org/10.1080/01425699995434>.
- Mason, Elinor, and Alan T. Wilson. 2017. "Vice, Blameworthiness, and Cultural Ignorance." In *Responsibility: The Epistemic Condition*, edited by Philip Robichaud and Jan Willem Wieland, 82–100. Oxford: Oxford University Press. <https://doi.org/10.1093/oso/9780198779667.001.0001>.
- Mason, Rebecca. 2011. "Two Kinds of Unknowing." *Hypatia* 26 (2): 294–307.
- McDowell, John. 1996. *Mind and World*. 2nd ed. Cambridge, MA: Harvard University Press.
- . 1998a. "Non-Cognitivism and Rule-Following." In *Mind, Value, and Reality*, 198–220. Cambridge, MA: Harvard University Press.
- . 1998b. "Virtue and Reason." In *Mind, Value, and Reality*, 50–73. Cambridge, MA: Harvard University Press.
- McEwan, Ian. 1998. *Enduring Love*. London: Vintage.
- McHugh, Conor. 2013. "Epistemic Responsibility and Doxastic Agency." *Philosophical Issues* 23: 132–57.
- Medina, José. 2013. *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and Resistant Imaginations*. Oxford: Oxford University Press.
- Michalko, Rod. 1998. *The Mystery of the Eye and the Shadow of Blindness*. Toronto, ON: University of Toronto Press.
- Midgley, Mary. 2006. "Philosophical Plumbing." In *Utopias, Dolphins, and Computers: Problems in Philosophical Plumbing*, 1–12. London: Routledge.
- Mill, John Stuart. 1979. *Utilitarianism*. Indianapolis, IN: Hackett.
- Mills, Charles W. 2007. "White Ignorance." In *Race and Epistemologies of Ignorance*, edited by Shannon Sullivan and Nancy Tuana, 13–57. Albany, NY: The State University of New York Press.
- Morvan, Pierre Le, and Rik Peels. 2016. "The Nature of Ignorance: Two Views." In *The Epistemic Dimensions of Ignorance*, edited by Rik Peels and Martijn Blaauw, 12–32. Cambridge, UK: Cambridge University Press.
- Müller, Jozef. 2015. "Aristotle on Vice." *British Journal for the History of Philosophy* 23 (3): 459–77.
- Murdoch, Iris. 1959. "The Sublime and the Good." *Chicago Review* 13 (3): 42–55.
- . 1970. *The Sovereignty of Good*. 2013 ed. London: Routledge.
- Nagel, Thomas. 1979. "Moral Luck." In *Mortal Questions*, 24–38. Cambridge, UK:

- Cambridge University Press. <https://doi.org/10.1017/CBO9781107341050.005>.
- . 1989. *The View from Nowhere*. Oxford, UK: Oxford University Press.
- Nelkin, Dana K. 2013. "Moral Luck." Edited by Edward N Zalta. *The Stanford Encyclopedia of Philosophy*.
- Neurath, Otto. 1973. "Anti-Spengler." In *Empiricism and Sociology*, 158–213. Dordrecht: D. Reidel Publishing Co.
- Nietzsche, Friedrich. 1998. *On the Genealogy of Morality*. Translated by Maudemarie Clark and Alan J. Swensen. Indianapolis, IN: Hackett.
- Orwell, George. 1949. *Nineteen Eighty-Four*. London: Martin Secker & Warburg, Ltd.
- Oshana, Marina A. L. 1997. "Ascriptions of Responsibility." *American Philosophical Quarterly* 34 (1): 71–83.
- Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Clarendon Press.
- Peacocke, Christopher. 1992. *A Study of Concepts*. Cambridge, MA: MIT Press.
- Peirce, Charles S. 1992a. "The Doctrine of Chances." In *The Essential Peirce: Selected Philosophical Writings, Volume 1 (1867–1893)*, edited by Nathan Houser and Christian Kloesel, 143–54. Bloomington and Indianapolis: Indiana University Press.
- . 1992b. "The Fixation of Belief." In *The Essential Peirce: Selected Philosophical Writings, Volume 1 (1867–1893)*, edited by Nathan Houser and Christian Kloesel, 109–23. Bloomington and Indianapolis: Indiana University Press.
- Phillips, Michael, and Trevor Phillips. 1998. *Windrush: The Irresistible Rise of Multi-Racial Britain*. London: Harper Collins.
- Pinker, Steven. 1989. *Learnability and Cognition: The Acquisition of Argument Structure*. Cambridge, MA: MIT Press.
- Plato. 2006. "Euthyphro." In *Classics of Western Philosophy*, edited by Steven M. Cahn, translated by G. M. A. Grube, 6th ed., 20–28. Indianapolis, IN: Hackett.
- Pohlhaus, Gaile. 2012. "Relational Knowing and Epistemic Injustice: Toward a Theory of Willful Hermeneutical Ignorance." *Hypatia* 27 (4): 715–35. <https://doi.org/10.1111/j.1527-2001.2011.01222.x>.
- Pratchett, Terry. 2008. *Nation*. New York: Doubleday.
- Pratt, Richard H. 1892. "The Advantages of Mingling Indians with Whites." In *Official Report of the Nineteenth Annual Conference of Charities and Correction*, edited by Isabel C Barrows, 46–59. Denver, CO: George H. Ellis.
- Price, Eleanor, and Christiane Wisheart. 2018. "28: Philosophy and #MeToo with Emily McWilliams." *Examining Ethics*. Podcast.
- Prinzing, Michael. 2017. "The Revisionist's Rubric: Conceptual Engineering and the Discontinuity Objection." *Inquiry: An Interdisciplinary Journal Of Philosophy* Online First.



- Putnam, Hilary. 1981. "Two Conceptions of Rationality." *Reason, Truth and History*, Cambridge, UK: Cambridge University Press. 103–26.
- Quine, Willard Van Orman. 1953. "Two Dogmas of Empiricism." In *From a Logical Point of View*. Cambridge, MA: Harvard University Press. 20–46.
- Railton, Peter. 1984. "Alienation, Consequentialism, and the Demands of Morality." *Philosophy and Public Affairs* 13: 134–71.
- Roberts, Debbie. 2013. "Thick Concepts." *Philosophy Compass* 8 (8): 677–88. <https://doi.org/10.1111/phc3.12055>.
- Robison, John Elder. 2013. "What Is Neurodiversity?" *Psychology Today*, October 7, 2013. <https://www.psychologytoday.com/us/blog/my-life-aspergers/201310/what-is-neurodiversity>.
- Romdenh-Romluc, Komarine. 2016. "Hermeneutical Injustice: Blood-Sports and the English Defence League." *Social Epistemology* 30 (5–6). Routledge: 592–610. <https://doi.org/10.1080/02691728.2016.1172363>.
- Rosch, Eleanor, and Carolyn Mervis. 1975. "Family Resemblances: Studies in the Internal Structure of Categories." *Cognitive Psychology* 7: 573–605.
- Rosen, Gideon. 2003. "Culpability and Ignorance." *Proceedings of the Aristotelian Society* 103: 61–84.
- . 2004. "Skepticism about Moral Responsibility." *Philosophical Perspectives* 18: 295–313.
- Santelli, John, Mary A. Ott, Maureen Lyon, Jennifer Rogers, Daniel Summers, and Rebecca Schleifer. 2006. "Abstinence and Abstinence-Only Education: A Review of U.S. Policies and Programs." *Journal of Adolescent Health* 38: 72–81.
- Scanlon, Thomas M. 1998. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
- Segrave, Kerry. 2014. *Beware the Masher: Sexual Harassment in American Public Places, 1880–1930*. Jefferson, NC: McFarland & Co.
- Seidman, Jeffrey. 2005. "Two Sides of 'Silencing.'" *The Philosophical Quarterly* 55 (218): 68–77.
- Sher, George. 2009. *Who Knew? Responsibility Without Awareness*. Oxford: Oxford University Press.
- Shoemaker, David. 2007. "Moral Address, Moral Responsibility, and the Boundaries of the Moral Community." *Ethics* 118: 70–108.
- . 2011. "Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility." *Ethics* 121 (3): 602–32. <https://doi.org/10.1086/659003>.
- . 2017. "Response-Dependent Responsibility; or, A Funny Thing Happened on the Way to Blame." *The Philosophical Review* 126 (4): 481–527.
- Shotwell, Alexis. 2011. *Knowing Otherwise: Race, Gender, and Implicit Understanding*.

University Park, PA: Penn State University Press.

- Shrader-Frechette, Kristin. 2010. "Conceptual Analysis and Special-Interest Science: Toxicology and the Case of Edward Calabrese." *Synthese* 177 (3): 449–69. <https://doi.org/10.1007/s11229-010-9792-5>.
- Sidgwick, Henry. 1981. *The Methods of Ethics*. 7th ed. Indianapolis, IN: Hackett.
- Smith, Angela M. 2000. "Identification and Responsibility." In *Moral Responsibility and Ontology*, edited by Ton van den Beld, 233–46. Dordrecht: Kluwer Academic Publishers.
- . 2005. "Responsibility for Attitudes: Activity and Passivity in Mental Life." *Ethics* 115 (2): 236–71. <https://doi.org/10.1086/426957>.
- . 2008. "Control, Responsibility, and Moral Assessment." *Philosophical Studies* 138 (3): 367–92.
- . 2012. "Attributability, Answerability, and Accountability: In Defense of a Unified Account." *Ethics* 122 (3): 575–89. <https://doi.org/10.1086/664752>.
- Smith, Edward E., and Douglas L. Medin. 1981. *Categories and Concepts*. Cambridge, MA: Harvard University Press.
- Smith, Holly M. 1983. "Culpable Ignorance." *The Philosophical Review* 92 (4): 543–71.
- Stanley, Jason. 2016. *How Propaganda Works*. Princeton, NJ: Princeton University Press.
- Stein, Joseph, Jerry Bock, and Sheldon Harnick. 1964. *Fiddler on the Roof: Based on Sholom Aleichem's Stories*. New York: Crown Publishers. [https://archive.org/details/fiddleronroofoobock\\_0](https://archive.org/details/fiddleronroofoobock_0).
- Stocker, Michael. 1976. "The Schizophrenia of Modern Ethical Theories." *The Journal of Philosophy* 73: 453–66.
- Strawson, Peter F. 2008. "Freedom and Resentment." In *Freedom and Resentment and Other Essays*, 1–28. Abingdon: Routledge.
- Taylor, Charles. 1976. "Responsibility for Self." In *The Identities of Persons*, edited by Amélie E. Rorty, 281–99. Berkeley, CA: University of California Press.
- The British Philosophical Association. n.d. "The BPA/SWIP Good Practice Scheme." Accessed June 28, 2018. <http://bpa.ac.uk/resources/women-in-philosophy/good-practice>.
- Thomas, Clarence. 2007. *My Grandfather's Son: A Memoir*. New York: Harper Perennial.
- Tilly, Charles. 2005. *Identities, Boundaries, and Social Ties*. Boulder, CO: Paradigm Publishers.
- Tilly, Charles, and Sidney Tarrow. 2006. *Contentious Politics*. Oxford: Oxford University Press.
- Toronto Mad Pride. 2017. "Mission Statement." Mad Times. 2017. <http://www.torontomadpride.com/issue/issue1/>.

- Townsend, Mark. 2018. "Transgender Woman Sues over Ordeal in Male Prison." *The Guardian*, January 20, 2018.  
<https://www.theguardian.com/society/2018/jan/20/tara-hudson-transgender-prisoner-sues-government>.
- Truth and Reconciliation Commission of Canada. 2015. "Honouring the Truth, Reconciling for the Future: Summary of the Final Report of the Truth and Reconciliation Commission of Canada." National Centre for Truth and Reconciliation, University of Manitoba. [http://nctr.ca/assets/reports/FinalReports/Executive\\_Summary\\_English\\_Web.pdf](http://nctr.ca/assets/reports/FinalReports/Executive_Summary_English_Web.pdf).
- Väyrynen, Pekka. 2009. "Objectionable Thick Concepts in Denials." *Philosophical Perspectives* 23 (Ethics): 439–69.
- . 2017. "Thick Ethical Concepts." Edited by Edward N Zalta. *The Stanford Encyclopedia of Philosophy*.  
<https://plato.stanford.edu/archives/fall2017/entries/thick-ethical-concepts/>.
- Watson, Gary. 1975. "Free Agency." *Journal of Philosophy* 72: 205–20.
- . 1996. "Two Faces of Responsibility." *Philosophical Topics* 24 (2): 227–48.  
<https://doi.org/10.5840/philtopics199624222>.
- Weiskopf, Daniel Aaron. 2009. "The Plurality of Concepts." *Synthese* 169: 145–73.
- White, Edmund. 1983. *A Boy's Own Story*. London: Picador.
- Wilde, Oscar. 1908. *The Picture of Dorian Gray*. Leipzig: Bernhard Tauchnitz.
- Wilde v. Queensberry. 1895. <http://www.famous-trials.com/wilde/330-libel>.
- Williams, Bernard. 1981. "Moral Luck." In *Moral Luck: Philosophical Papers 1973-1980*, 20–39. Cambridge, UK: Cambridge University Press.
- . 1985. *Ethics and the Limits of Philosophy*. London: Fontana Press.
- . 1993. "Postscript." In *Moral Luck*, edited by Daniel Statman, 251–58. Albany, NY: SUNY Press.
- . 1995. "Truth in Ethics." *Ratio* 8 (3): 227–42.
- . 2002. *Truth and Truthfulness: An Essay in Genealogy*. Princeton, NJ: Princeton University Press.
- Wittgenstein, Ludwig. 1978. *Philosophical Investigations*. Oxford: Basil Blackwell.
- Wolf, Susan. 1988. "Sanity and the Metaphysics of Responsibility." In *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*, edited by Ferdinand David Schoeman, 46–62. Cambridge, UK: Cambridge University Press.
- . 2001. "The Moral of Moral Luck." *Philosophic Exchange* 31 (1): Article 1.  
[https://digitalcommons.brockport.edu/phil\\_ex/vol31/iss1/1/](https://digitalcommons.brockport.edu/phil_ex/vol31/iss1/1/).
- Zack, Naomi. 1999. "White Ideas." In *Whiteness: Feminist Philosophical Reflections*, edited by Chris J Cuomo and Kim Q Hall, 77–84. Lanham, MD: Rowman & Littlefield.

Zimmerman, Michael J. 1988. *An Essay on Moral Responsibility*. Totowa, NJ: Rowman & Littlefield.

———. 1997. "Moral Responsibility and Ignorance." *Ethics* 107 (3): 410–26.