DEMYSTIFYING NORMATIVITY

MORALITY, ERROR THEORY, AND THE AUTHORITY OF NORMS

Eline Gerritsen

This PhD project was financially supported by:
Prins Bernhard Cultuurfonds
Stichting dr. Hendrik Muller's Vaderlandsch Fonds
University of Groningen

Cover photo: Paul Earle
Layout: Jakob Hinze
Printing: Ipskamp Printing

© 2022 Eline Gerritsen

All rights reserved. No part of this publication may be reproduced or distributed in any form or by any means without written permission from the author.

Demystifying Normativity

Morality, Error Theory, and the Authority of Norms

PhD thesis

to obtain the degree of PhD of the
University of Groningen
on the authority of the
Rector Magnificus Prof. C. Wijmenga
and in accordance with
the decision by the College of Deans

and

to obtain the degree of PhD of the
University of St Andrews
on the authority of the
Chancellor The Rt Hon Lord Campbell of Pittenweem

and

to obtain the degree of PhD of the
University of Stirling
on the authority of the
Chancellor The Rt Hon Lord McConnell of Glenscorrodale

Joint PhD degree

This thesis will be defended in public on

Monday 29 August 2022 at 16.15 hours

by

Eline Gerritsen

born on 31 March 1995
in Winterswijk

Supervisors

Dr. K. Hurtig
Prof. B. Streumer

Co-supervisors

Dr. H.W.A. Evers
Dr. J. Snedegar

Assessment Committee

Prof. R. Cruft
Prof. F.A. Hindriks
Prof. J. Olson
Prof. C. Tiefensee

Thesis Abstract

We are subject to many different norms telling us how to act, from moral norms to etiquette rules and the law. While some norms may simply be ignored, we live under the impression that others matter for what we ought to do. How can we make sense of this normative authority some norms have? Does it fit into our naturalist worldview? Many philosophers claim it does not. Normativity is conceived to be distinct from ordinary natural properties, making it mysterious. The mystery fuels a radical yet prominent scepticism about the existence of normative properties: if they are too strange to actually exist, there is nothing we ought to do. Some take this to mean, moreover, that nothing is morally right or wrong. We must critically examine the ideas behind these theories, which put both morality and normativity in general on the line.

The aim of this thesis is to unravel the mystery of normativity. It uncovers and objects to the influential non-natural conception of it, arguing that we can capture normativity with natural properties. In particular, it explores how the authority of norms can be explained by the commitments of the people subject to them. In connection to this, it challenges the conceptual claim behind the view that all moral judgements are mistaken. Finally, it reveals that treating morality as a mere fiction has revolutionary practical implications. This emphasises the importance of the overall conclusion: we need not conceive of either moral or normative properties as too mysterious to exist.

Contents

Acknowledgements ..... *xi*

Introduction *1*

1. The puzzle of normativity 1
2. Thesis summary 4

CHAPTER 1
Prichard's Mistake: In Defence of 'Why Be Moral?' *9*

0. Introduction 9
1. Moral content versus normativity 11
2. Scanlon's dilemma 16
3. The self-evidence of reasons to be moral 22
4. Conclusion 29

CHAPTER 2
Demystifying Authoritative Normativity *31*

0. Introduction 31
1. Two kinds of normativity 32
2. Defining authoritative normativity 37

3. Ought *simpliciter* 42

4. Irreducibility 48

4.1 Irreducibility and norms 49

4.2 Irreducible authority 53

4.3 Categorising authority 58

5. Conclusion 62

CHAPTER 3

Sources of Authority *65*

0. Introduction 65

1. Forms of justification 67

1.1 A norm-based account of authoritative normativity 67

1.2 Intrinsic justification 68

1.3 Coherence 70

1.4 Foundations of justification 73

2. Explaining authority 78

2.1 Practical identity 78

2.2 The practical standpoint 82

2.3 Desires 88

2.4 Collective flourishing and constitutive aims 90

3. Conclusion 94

CHAPTER 4

Conventions, Contingency, and Crossing the Is/Ought Gap *99*

0. Introduction 99

1. Implications of a Humean account of authority 101

1.1 The theory in practice 101
1.2 The authority of conventional norms 104
2. The problem of contingency 109
2.1 Contingency and the concept of authority 110
2.2 Implications of contingency 113
3. Pulling the rabbit out of the hat 122
3.1 The is/ought gap 122
3.2 The authority of the means-end principle 127
4. Conclusion 131

CHAPTER 5
The Conceptual Thesis of Moral Error Theory *133*
0. Introduction 133
1. Mackie's conceptual thesis 135
2. Categorical requirements and authority 138
3. Olson's conceptual thesis 141
4. Joyce's conceptual thesis 145
5. Rejecting the conceptual thesis 148
5.1 Categoricity 148
5.2 Authority 151
5.3 Categorical authority 153
6. Conclusion 160

CHAPTER 6
The Second Revolution of Moral Fictionalism *161*
0. Introduction 161

1. The rationale for revolutionary moral fictionalism 163

2. Fictionalism and contractarianism 169

3. Objections and replies 175

3.1 The feasibility of changing moral thinking 175

3.2 Other-regarding preferences 176

3.3 So what? 179

4. Conclusion 181

Conclusion *183*

Nederlandse samenvatting *189*

Curriculum Vitae *195*

Bibliography *197*

Acknowledgements

This thesis is the result of four years of research, at three universities located in two different countries, during one global pandemic. It will not come as a surprise that I have a few people to thank for helping me to get to the end.

Vital to the success of any PhD project are its supervisors. I have been lucky to have a gradually expanding team of excellent supervisors behind me, who were always willing to give me any support I needed. Kent Hurtig has supervised me since the MLitt module that first drew my attention to questions about normativity. Over the years, he has helped me to narrow my focus and get a better grasp of the complex issues involved, always with positive encouragement. He has been dedicated to my progress, spending entire afternoons discussing every detail of my work. After I visited Groningen for a semester to work with Bart Streumer, Bart invited me to join the department for the rest of my PhD and agreed to become my second primary supervisor. This turned out to be an excellent opportunity, for which I am grateful. Bart has had a very positive effect on the quality of my research – draft after draft, he showed me how to improve my arguments, even (or especially) those against his own views. Daan Evers joined nearly all of our meetings and gave me helpful feedback on each chapter, going far beyond what might be expected of a secondary supervisor. Justin Snedegar offered in-depth discussions of my work and many insightful suggestions, both before and after he became my other secondary supervisor.

I was honoured to have Sarah Broadie as my secondary supervisor for the first years of my PhD, which she likely agreed to more out of kindness than out of a personal interest in metaethics. As intimidated and inspired as I was by her brilliant mind, she

always took my ideas seriously. She guided me in shaping my project in the early stages and, as I hoped she would, helped me to keep the timeless philosophical questions behind my research in focus. Sarah's recent passing is a great loss to philosophy.

I want to thank my annual reviewers – Rowan Cruft, Adam Etinson, James Harris, and Ben Sachs – for useful conversations about my thesis as well as my general academic development. My thesis has also benefitted from generous feedback by Rachel Handley, Erik Kassenberg, Jonas Olson, Janis Schaab, Merel Semeijn, Ravi Thakral, Lizzy Ventham, and participants of the Ethics, Social and Political Philosophy seminar at Groningen.

This research was supported by PhD scholarships from Prins Bernhard Cultuurfonds, Hendrik Muller Fonds and the University of Groningen. The Scottish Graduate School for Arts & Humanities funded my doctoral internship at the Carnegie Trust for the Universities of Scotland. I received an Erasmus+ grant twice: the first funded a semester at St Andrews during my Bachelor, while the second allowed me to visit Groningen during my PhD. Both were crucial steps in the academic journey that led to the completion of this PhD.

St Andrews has been an excellent academic home in which to grow from a student into a researcher. I have profited from the stimulating intellectual environment in Edgecliffe as well as the friendly solidarity among fellow PhD students. At Groningen, I especially enjoyed being a part of the metaethics reading group, which produced many interesting discussions and grew my confidence in my own expertise.

As undertaking a philosophy PhD in a remote Scottish town and in times of social distancing can be a particularly isolating experience, it was invaluable to have long-distance friendships I could continuously count on. I am particularly grateful to the lovely community of philosophers I have known for the past years – Jaana, Jade, Joe, Justin, Katherine, Leonie, Liv, Rebeccas, Sara, and others already mentioned – for answering my many questions about academic matters and providing a constant stream of sympathy, encouragement, and cat pictures.

I am grateful to my parents for always supporting me in my unusual career choices, even when they take me far away from them, just because I say I have to make them. Finally, I owe my deepest thanks to Jakob Hinze – for proofreading and formatting, for inspiring me as an exemplar of a hard worker and good researcher, for keeping me

company during multiple lockdowns and international moves, but most importantly, for making me happy enough to be able to write this thesis.

Introduction

1. The puzzle of normativity

We live under the impression that there are actions we ought to perform. We ought to be kind, help those in need, and refrain from breaking our promises. We also ought to look before crossing a street, keep an appropriate distance to others in a queue, and refrain from insulting our host. Our behaviour is constantly subject to requirements of different kinds, which often shape what we actually do. As common as this experience of normativity is, it poses a philosophical mystery: where do these 'oughts' come from? How do they fit into the world? Actions come with various features that are not particularly mysterious – we can give a relatively straightforward explanation of what a certain action consists in, what effects it is likely to have, and whether and why someone desires to do it. It seems, however, that actions can also have the more puzzling property of being what we ought to do. Similarly, they can be required, permissible, or impermissible. We can have a reason to do certain things and not others. How can we make sense of such normative properties?

The normativity of morality, in particular, has been the subject of much philosophical debate. The common idea that we are bound to morality's demands in an important sense sparks the question how we can account for this. However, the puzzle of normativity is not exclusive to moral norms. There is a large variety of norms beyond moral ones that tell us what to do – for example, the law demands that I stop at a stop sign and prudence requires me to choose the action that is in my long-term best interest. At least some of these prescriptions appear to have a significant normative

force as well, such that I am not free to ignore them. This normative force calls for an explanation just as the normativity of morality does.

In the current metanormative literature, attempts to solve this puzzle have taken an undesirable direction: normativity tends to be presented as something mysterious which simply cannot be unravelled. A satisfying account of normative properties would explain them in terms of natural properties we are familiar with; yet, it is typically treated as impossible to capture the normative force we are interested in in this way. As a result, we appear to have two main options, which both receive substantial support. The first option is endorsing non-naturalism: we can agree that normative properties are mysterious, in the sense of being unlike natural properties, but accept that such mysterious properties exist. This comes with serious metaphysical and epistemological costs and, in a way, leaves the nature and practical relevance of normative properties unexplained. The second option is to settle for a normative error theory: we can agree that normative properties would have to be mysterious and conclude from this that they do not exist. Since error theory claims that our normative judgements do ascribe such properties, it entails that all positive normative judgements are mistaken. We do not, in fact, have a real reason to do anything. This is a radical theory, which gives up on the normativity we are trying to explain. In addition, it forms the basis of an influential form of *moral* error theory: when having a significant normative force is considered to be an essential feature of moral properties, scepticism about normativity leads to the view that nothing is morally wrong or right, contradicting some of our strongest intuitions. Both non-naturalism and error theory, therefore, come with unpalatable implications.

I provide an alternative to these prominent theories by showing how we can take the mystery out of normativity. While the common understanding of normative properties implies that they occupy their own special domain, separate from the natural world, I take them back to earth. The main research question of this thesis, which is the explicit subject of some of its chapters and forms the background of others, is this: how can we make sense of normativity in a way that does not lead to an error theory? In contrast to non-naturalists, my strategy is not to make room for the mystery, but to unravel it: I question error theory's conceptual assumptions about normativity. I argue that we need not conceive of normativity as something that is

plausibly too strange to exist – there is room to understand it in a way that makes normative force perfectly metaphysically and epistemologically acceptable.[1]

My focus is largely on metaphysically naturalist theories of normativity as an alternative to error theory. By 'metaphysically naturalist', I mean theories that account for normativity without positing any non-natural properties. My discussion also covers theories not labelled 'naturalism', including both Kantian and Humean constructivism. The metanormative divide of interest here is not between realism and anti-realism, but between theories with a non-naturalist understanding of normativity – including both non-naturalist realism and error theory – and theories that interpret normativity in such a way that its instantiation does not require any non-natural, irreducible normative properties. I explore, in particular, how moral and normative error theory can be countered by theories that explain normativity by appealing to agents' contingent commitments, such as desires, aims or normative judgements. My discussion will not cover non-cognitivist interpretations of normative claims; it carries a background assumption that cognitivism about normative language is correct.[2]

An important thread running through the thesis is the increasingly recognised distinction between two forms of normativity, which I refer to as formal and authoritative normativity. Formal normativity is the normativity all norms have simply in virtue of requiring actions or attitudes. Authoritative normativity involves a more significant force – it concerns what we *really* have reason to do, in a way we cannot legitimately ignore. We ascribe this to some norms, but not all: consider the normative difference between moral norms and an entirely arbitrary set of rules for life I have made up. Authoritatively normative demands on us matter. The puzzle of normativity I am interested in is specifically a puzzle of authoritative normativity: this form of normativity is both more important and more difficult to make sense of in natural terms.

My discussion puts much weight on the role of norms in normativity. I use a norms-first account of normativity, contrasted, for example, with a reasons-first alternative. I treat the puzzle of authoritative normativity as a question of how, and why, some norms have authority. This is not currently the most standard approach; instead,

---

[1] I use 'strange' in place of the term 'queer', which metaethicists traditionally use in this context, because the latter is unnecessarily loaded.

[2] It will also not cover non-metaphysical cognitivism or cognitive irrealism, since this would require an elaborate discussion of metaphysics and truth which would turn this into a different thesis.

I have taken inspiration from the debates on normativity that took place around the turn of the millennium.[3] In my view, this focus on the authority of norms allows us to successfully capture the core problem of normativity, as well as the potential solutions to consider. I do not claim that a norm-based approach is the only or best way of making sense of authoritative normativity. Instead, what I offer is a way of thinking about normativity which I have personally found highly beneficial for getting to grips with the topic. My expectation, and hope, is that it can elucidate the seemingly mysterious notion of authoritative normativity for others as well.

2. Thesis summary

Chapter 1: Prichard's mistake: In defence of 'why be moral?'

While chapters two, three and four can be seen to form one large argument concerning authoritative normativity together, chapters one, five and six tackle different questions with the shared theme of moral normativity. The first chapter responds to H.A. Prichard's claim that 'why be moral?' is a mistaken question. I discuss and reject the thought that no appropriate reason can be given, as well as the idea that it is self-evident that we have a reason to be moral. I argue that, because we should understand it to ask for authoritative reasons, 'why be moral?' is meaningful and open-ended after all. This chapter functions, at the same time, as an introduction to the topic of the authority of norms and moral normativity in particular: I explain how questions about the content of morality relate to questions about its normative force and showcase a range of different views on the relation between morality and normativity.

[3] Readers will encounter these sources of inspiration in my references – think of Christine Korsgaard, *The Sources of Normativity* (Cambridge: Cambridge University Press, 1996); Jean Hampton, *The Authority of Reason* (Cambridge: Cambridge University Press, 1998); David Copp, *Morality, Normativity, and Society* (Oxford: Oxford University Press, 2001).

Chapter 2: Demystifying authoritative normativity

In chapter two, I analyse the concept of authoritative normativity. I explain the distinction between formal and authoritative normativity and show the various ways in which this contrast has been characterised in the literature. I argue that common ways of capturing authoritative normativity are not successful: defining features that have been ascribed to it, including prescriptivity and being reason-giving, can also be ascribed to formal normativity. Next, I separate the concept of authoritative normativity from that of ought *simpliciter* and show how these relate to each other. In the rest of the chapter, I argue against the popular view that authoritative normativity is, by conceptual necessity, irreducible and intrinsic to the norms that have it. This view is central to Jonas Olson's normative error theory – it interprets normative authority as a non-natural force that seems too strange to exist. Yet, we must not take for granted that authoritative normativity is mysterious in this way; I argue that there is conceptual room for understanding it to be explained by natural facts about agents.

Chapter 3: Sources of authority

The third chapter elaborates on which interpretations of authoritative normativity are available beyond the standard view that it is an irreducible and intrinsic property. I first offer an abstract norm-based account according to which normativity is authoritative if and only if it is grounded in justified norms. This provides a framework for categorising more specific theories of authoritative normativity in accordance with their view on what justifies norms. When we recognise these different categories, it becomes clear that appealing to a mysterious intrinsic justification is far from the only way of making sense of normative authority. Most of the chapter is focused on exploring a range of metaphysically naturalist theories – Kantian and Humean constructivism, desire-based views, as well as theories that account for norms' authority in terms of collective flourishing. By showing how the special force of authoritative normativity could be explained by our commitments, these theories offer an important alternative to the non-naturalist conception of authority that is behind normative error theory.

Chapter 4: Conventions, contingency, and crossing the is/ought gap

In chapter four, I narrow my focus to Humean theories of authoritative normativity, which explain it in terms of facts about agents' contingent commitments. This includes both Humean constructivism and desire-based views. First, I explain what such views imply about the authority of different kinds of norms for agents with particular commitments. An interesting upshot is that conventional norms can have authority on Humean views – this, I argue, speaks in favour of them. The rest of the chapter is dedicated to defending Humean theories against the most striking objections to them. I argue that the counterintuitive contingency involved in them can be minimised and does not constitute an unacceptable cost, especially when we consider the problems of non-naturalist alternatives. A different major concern is that Humean theories attempt to cross the is/ought gap. In response, I question this gap and show that the idea that the normative is fundamentally different from the natural presupposes a non-natural interpretation of authority. Finally, I explain how the Humean can account for the authority of the means-end principle, which takes an important role in their theory.

Chapter 5: The conceptual thesis of moral error theory

In chapter five, I critically assess a prominent normativity-focused form of moral error theory. Arguments for a moral error theory consist of a conceptual thesis, which ascribes a certain non-negotiable feature to ordinary moral concepts, and a substantive thesis, which denies that this feature exists. I counter error theory by calling the conceptual thesis into question. I start by discussing how we should understand the conceptual thesis of Mackie, Olson, and Joyce. In my terms, Olson and Joyce ascribe a categorical form of authoritative normativity to morality. Next, I object to this conceptual thesis. I argue that it is not obvious that morality has categorical authority; while it is plausible to ascribe categoricity to morality, this only establishes that moral norms apply categorically. Olson and Joyce need more than this; their problem is not with categorical formal normativity, but with categorical authoritative normativity. Yet, it is not clear that ordinary users of moral concepts have a commitment, let alone a conceptual commitment, to us having desire-independent authoritative reasons to

be moral. I argue that it is an open question whether morality is necessarily authoritative for everyone – therefore, it is not the case that there can be no moral properties if there is no categorical authoritative normativity.

Chapter 6: The second revolution of moral fictionalism

The final chapter takes the topic of our reasons to be moral in a different direction: it asks what we should do with moral discourse and practice if moral error theory is correct. According to Joyce's revolutionary moral fictionalism, error theorists should pretend that there are true moral propositions in order to keep the benefits moral thinking has for their preference-satisfaction. This, he claims, frees error theory from radical practical implications. In response, I argue that implementing fictionalism would not preserve our moral practices, but disrupt them. The change from moral belief to make-belief yields an unintended second revolution: a revolution in the content of morality. I show that fictionalism necessarily relies on a similar justification of moral practices as David Gauthier's contractarianism, and consequently has similar implications for moral content. Because fictionalists engage in moral thinking purely for its instrumental value, they should only accept moral obligations that are useful to them into their fiction. This restriction is important: I argue that the most useful moral fiction departs substantially from conventional moral views. Revolutionary moral fictionalism is therefore more radical than it is promised to be.

CHAPTER 1

Prichard's Mistake: In Defence of 'Why Be Moral?'

0. Introduction

'Why be moral?' is one of the classic questions of moral philosophy. No matter how certain we are that particular actions are morally required, an amoral person can come along and ask, 'What is that to me? Why must I obey these moral requirements instead of doing what I feel like?' Philosophers have tried to answer the amoralist at least since Plato's Glaucon and Adeimantus challenged Socrates to explain why it is better for a person to live a just life.[1] The topic has been approached with a wide variety of strategies to show that we do have a reason to do what is morally required, that we do not, or that morality holds an important power over most of us even if we cannot convince the amoral person. There is also a range of interpretations on how significant the question is and what is at stake if no adequate justification for being moral can be given. What all treatments of 'Why be moral?' do have in common is the assumption that there is a point to concerning oneself with this question. This is what is famously

[1] Plato, *Republic*, trans. C. D. C. Reeve (Indianapolis: Hackett Publishing Company, 2004), Book Two.

denied by H.A. Prichard in "Does Moral Philosophy Rest on a Mistake?". In this paper, he bluntly expresses his great dissatisfaction with all work done in the subject. The underlying problem, as diagnosed by Prichard, is not a minor one: he suggests that moral philosophy "consists in the attempt to answer an improper question." Therefore, "the existence of the whole subject, as usually understood, rests on a mistake".

To justify an investigation into normativity with a special focus on the authority of morality, over a century after Prichard objected to this topic so strongly, I must show Prichard himself to be mistaken. In this chapter, I will explain and respond to Prichard's challenge. My aim is not only to argue that 'Why be moral?' is a meaningful question worthy of attention after all; I will also use this discussion to distinguish different positions on the relation between morality and normativity. My focus here is on opening up and clarifying the question of moral normativity – in the chapters that follow, I will elaborate on different conceptions of our reasons to be moral, as well as reasons to comply with other types of norms.

I will start with a discussion of what the fundamental question of moral philosophy is that Prichard sees as mistaken. In this context, I will clarify the difference between questions about moral content and questions about moral normativity. I will also introduce a classification of different types of ethical internalism. In the second section, I will discuss what is supposed to be wrong with 'Why be moral?', covering my own interpretation of Prichard's view as well as T.M. Scanlon's. The problem Scanlon identifies for 'Why be moral?' is that there is no satisfactory answer to it: either we say we have moral reasons to be moral, which is trivial, or we give non-moral reasons, which would be of the wrong kind. I will reject this dilemma by introducing a distinction between formal and authoritative reasons, arguing that 'Why be moral?' asks for the latter. As I will show, this means that it is not trivial to claim that we have moral reasons to be moral. In the third section, I will object to Prichard's view that it is self-evident that we have a reason to do what morality requires. To do this, I will reject different ways one could arrive at this conclusion, including the idea that it is contained in our moral concepts that we have a reason to do what is morally required. I will conclude that the answer to 'Why be moral?' is not self-evident, and that there is plenty of room left for a meaningful investigation of whether we have authoritative reasons to do what morality requires.

1. Moral content versus normativity

To see whether Prichard's frustration with moral philosophy is justified, we first need to get a good grasp of the fundamental mistake he ascribes to it. There are different plausible interpretations of this. As Prichard characterises it, the central question of moral philosophy comes up in a context where an agent realises that a particular action they believe to be morally required is not an action that is in their own interest. Imagine someone made a promise to a colleague to proofread her paper tonight, but as her friends are going to the cinema, staying home to read the paper means missing out on an entertaining evening. Faced with such a sacrifice, one may ask: "Is there really a reason why I should act in the ways in which hitherto I have thought I ought to act? May I not have been all the time under an illusion in so thinking?"[2] We are normally convinced that we are under an obligation to keep our promises, for example, but when doing so is disadvantageous, a convenient scepticism can arise. There are two different aspects of moral obligations this scepticism could be about. The first kind of scepticism is about the *content* of morality. Here, the question can be reformulated as follows: 'Does morality really require me to act in the ways I thought it did?' Does morality really require me to fulfil my promises? If we picture morality to have the form of a rulebook, being sceptical about the content of morality means questioning our preconceived ideas of what the rules in the book are.[3] If we take this scepticism to its extreme, we also ask whether there are any rules in it at all.

The second kind of scepticism, on the other hand, is about the *normativity* of morality. On this interpretation, the central question Prichard ascribes to moral philosophy can be reformulated as 'Is there really a reason to act in the ways morality requires?' Here, Prichard's assumption would be that we are normally convinced that if morality requires us to do a certain action, we have a reason to do it. But when moral demands get in our way, we suddenly wonder if they genuinely have this normative force, or if we are free to ignore them. In our example, the reluctant promise-keeper would thus think: 'I know that I am morally required to keep my promise, but do I

[2] H.A. Prichard, 'Does Moral Philosophy Rest on a Mistake?', *Mind* 21, no. 81 (1912): 21.

[3] The analogy between morality and a rulebook is not meant to presume the truth of any particular first-order moral theory; the prescription(s) in the book could take a variety of forms, such as 'Do not lie', 'Do whatever maximises the good', 'Be courageous', et cetera.

really have a reason to keep my promise?' This kind of sceptic is not investigating whether the moral rulebook contains the rules we think it does, but whether the rules play the role in our normative lives which we ascribe to them. The sceptic takes the presumed moral rules for granted, and wonders whether they really determine or track what we have reason to do. If this is the kind of scepticism Prichard intends to point to, then it is the question of what reason we have to be moral that Prichard treats as the central question of moral philosophy.

Prichard's discussion in "Does Moral Philosophy Rest on a Mistake?" does not provide conclusive evidence for whether he means to reject the question about moral content or the one about moral normativity. This is partly because of his ambiguous formulations of the issue, which sometimes seem to switch between the two topics, and partly because the terms that are natural to use for these matters are ambiguous themselves. 'Ought' is a prominent example. 'I ought to fulfil my promises' can mean two things: (1) a norm requires that I fulfil my promises. Here we can assume that this is a norm of morality, unless the context specifies a different kind of norm. On this understanding, what is claimed is that it is written in the rulebook of morality that it is correct to fulfil my promises.[4] An alternative plausible meaning is (2) fulfilling my promises is what I have most reason to do. On this understanding, the claim is an answer to a general 'What ought I to do?' question rather than to 'What ought I to do morally speaking?'.

Because ought-claims can be about either the content of norms or their normativity, it is not clear which aspect of morality Prichard has in mind when he says that what we normally demand from moral philosophers[5] is "to convince us that we really ought to do what in our non-reflective consciousness we have hitherto believed we ought to do, or, if not, to tell us what, if any, are the other things which we really ought to do, and to prove to us that he is right."[6] The same goes for passages about whether we have the obligations we usually think we do: 'I am obligated to fulfil my promises'

[4] Again, this rulebook analogy is not meant to preclude any views on what form the content of the moral rulebook takes. For example, we could imagine that it contains the rule 'Fulfil promises', or 'A is to fulfil her promise in situation S', or a more general principle from which it somehow follows that A is to fulfil her promise.

[5] To be more precise, in this section Prichard writes that this is what moral philosophers want Aristotle to do, and why they may be dissatisfied by his work.

[6] Prichard, 'Does Moral Philosophy Rest on a Mistake?', 33.

could be a claim about what morality requires, but could also mean that fulfilling my promises is what I ought to do, overall. Prichard does not clearly lead us to either interpretation.[7] One further example of ambiguous formulations is when he writes that, in our moment of scepticism, we "wonder whether we really ought to do what we usually call our duty"[8] and want to find evidence "that we ought to act in the ways usually called moral"[9]. It is possible to interpret him as fixating either on the content or normativity of morality here, depending on which part of the sentence we emphasise. First, if we focus on 'usually called moral' (or 'what we usually call our duty'), the question seems to be about moral content – are the actions we usually call 'moral' really moral? Second, if we focus on 'whether we really ought to do' (or 'whether we ought to act'), the question seems to be about our reasons to do the presumably morally required actions.

At this point, some may ask: are the questions about moral content and normativity actually separate? Even if they do not have the same meaning, are their answers not linked? The thought here is that if you have established that the actions we usually believe to be morally required are indeed morally required – that is, if you have dispelled scepticism about moral content – you have at the same time made clear that we do indeed have a reason to do the actions we usually believe to be morally required. In the example where someone is reluctant to do what she promised, this would mean that once she has realised that morality really does demand her to fulfil her promise, she must also realise that she has a reason to fulfil her promise.[10] The normative status of this action is no longer an open question. If this is the case, it is not very interesting whether Prichard is rejecting the question 'is what we believe to be morally required

[7] Korsgaard also points out this ambiguity in Prichard's discussion in Korsgaard, *The Sources of Normativity*, 39.

[8] Prichard, 'Does Moral Philosophy Rest on a Mistake?', 22.

[9] Prichard, 22.

[10] Whether she ought to do it or merely has a reason to, which may be outweighed, depends on the type of link between morality and normativity one assumes. Prichard is very strict in this respect and assumes that it is always the case that what is morally required is what you ought to do overall. This is one of the elements of his position we can plausibly question.

indeed morally required?' or 'do we indeed have a reason to do what we believe to be morally required?', as they are answered jointly.[11]

Behind this thought about the relation of the two questions lies the assumption that *ethical internalism* is true. As 'internalism' and even 'ethical internalism' is used for a confusingly broad variety of positions, we need to make a number of specifications to get to the position that is relevant for understanding Prichard's complaint. Ethical internalism, which I will henceforth refer to as 'internalism', is roughly the view that morality necessarily or always involves a practical dimension. This practical dimension can relate either to normative force or to motivational power: according to *reasons internalism*, morality involves normative reasons to do certain actions; according to *motivational internalism*, morality involves agents being motivated to do certain actions.[12]

To make these positions still less vague, we need to specify the location of this normative or motivational power of morality. The two main contenders are moral facts themselves and judgements about them. If moral facts themselves have normative or motivational power, as *existence internalism* says, then the mere fact that φ-ing is morally required either entails that I have a normative reason to φ, or that I am motivated to φ. If *judgement internalism* is true, then I am motivated or have a normative reason to φ whenever I believe φ-ing to be morally required, whether that belief is correct or not.[13] Although this is not an exhaustive list of the views captured by 'ethical internalism',[14] we have already established four separate views about the practical force of morality which we can independently accept or reject. The strongest combinations seem to be *reasons existence internalism* and *motivational judgement internalism*, since it is implausible that false beliefs about moral facts give us reasons for action, and likewise

[11] An answer to the latter question does not seem to provide an answer to the former question unless you assume that all normative reasons are moral reasons, like Prichard does.

[12] The terminology in this section is taken from Russ Shafer-Landau, *Moral Realism: A Defence* (New York: Oxford University Press, 2005), 142–45.

[13] Existence internalism is also known as agent internalism. Judgement internalism is sometimes called appraisal internalism, e.g. by David O. Brink, *Moral Realism and the Foundations of Ethics* (Cambridge: Cambridge University Press, 1989), 40.

[14] Other views include hybrid versions of the views I have mentioned. For example, *knowledge internalism* ascribes normative or motivational power to true moral judgements. A hybrid of a different kind is Michael Smith's view that moral judgements come with reasons to be motivated to perform an action (Michael Smith, *The Moral Problem* (Oxford: Blackwell, 1995)).

implausible that moral facts will motivate us even if we are not aware of them. For all positions, we can distinguish a weak and a strong version. The weak version says that the reasons or motivation that arise are defeasible or can be overridden, while the strong version denies this.[15] Thus, only *strong reasons existence internalism* entails that if one is morally required to φ, one ought to φ.

Finally, we should make an important but often overlooked distinction, highlighted by Jeffrey Goldsworthy, between *substantial internalism* and *conceptual internalism*.[16] Substantial internalism is the view that morality *actually* has practical force, understood as either normative or motivational power. Conceptual internalism, in contrast, is the view that it is *conceptually* true that morality has this practical force.[17] According to *conceptual reasons existence internalism*, it is part of the meaning of 'φ-ing is morally required' that one has a reason to φ. As I will not focus on motivational power or moral judgements from here on, I will simply refer to this view as 'conceptual internalism'.

Only if conceptual internalism is true can we take the question about what is morally required and the question of what we have reason to do as one. Separating the multitude of internalisms clarifies that conceptual internalism is a specific thesis about the meaning of our moral concepts which can be rejected even by someone who does believe that we often or always have a strong reason to be moral. I will express my doubts about its truth below.

This paper does not make fully clear whether Prichard assumes conceptual internalism to be true, although his intertwined discussion of moral content and normativity may be taken to suggest that he does. More explicit support can be found in his "Manuscript on Morals":

> There remains to be noticed a question about moral obligation which when stated nakedly looks so ridiculous that it seems a waste of time

[15] Shafer-Landau, *Moral Realism*, 143.

[16] Jeffrey Goldsworthy, 'Externalism, Internalism and Moral Scepticism', *Australasian Journal of Philosophy* 70, no. 1 (1992): 40–60.

[17] Goldsworthy distinguishes a third category, sociological internalism: this is the thesis that substantial internalism is generally believed to be true. (Goldsworthy, 'Externalism, Internalism and Moral Scepticism', p. 41.)

> even to refer to it. This is the question, 'Why are we bound, or why ought we, to do what is right?'. The question not only looks but is ridiculous. For to refer to a certain group of actions as right actions is to imply that we already know that they are right, and therefore, since 'right' is after all only a synonym for 'ought to be done', that we already know that we ought to do them.[18]

Here, Prichard clearly expresses a commitment to a strong version of conceptual internalism: he claims that knowing that an action is morally right is sufficient for knowing not only that we have a reason to do it, but even that we ought to do it, since this is part of the meaning of 'right'.[19]

While I have suggested that Prichard can be understood to target both the question about content and the question about normativity in "Does Moral Philosophy Rest on a Mistake?", the standard interpretation is that he sees the latter as the mistaken question responsible for the failure of moral philosophy. This interpretation is supported by Prichard's focus on justifications for being moral in his other works, as the quote above illustrates – he hardly conceals his contempt for 'Why be moral?' here. I will now move on to discuss what it is about this question that is supposed to make it mistaken, and challenge Prichard's damning verdict.

2. Scanlon's dilemma

Why is it 'a waste of time' to ask whether or why we ought to do what is morally required, according to Prichard? He does not deny that we sometimes wonder whether we really ought to do something deemed obligatory, particularly when doing this would be detrimental to us. He also does not seem to judge those who ask this

[18] H. A. Prichard, 'Manuscript on Morals', in *Moral Writings*, ed. Jim MacAdam (Oxford: Oxford University Press, 2002), 119.

[19] Prichard even takes this a step further than conceptual internalism. He claims not only that 'ought to be done' is entailed by 'is right', but that these terms are synonymous. This means that we ought to do something if and *only* if it is morally right. This follows from Prichard's view of reasons, which I discuss below.

question in this spontaneous way as mistaken. Instead, in Prichard's eyes, the mistake lies in the way philosophers take it to heart as a serious and open question that needs to be answered with their usual tool, arguments. He believes it is a waste of time for a moral philosopher to investigate whether we ought to keep our promises, for example, because the careful reflection they would apply to this question is neither required nor suitable to answer it: "we do not come to appreciate an obligation by an *argument, i.e.* by a process of non-moral thinking".[20] Instead, "[t]he sense that we ought to do certain things arises in our unreflective consciousness, being an activity of moral thinking occasioned by the various situations in which we find ourselves."[21] Prichard believes that it is normally entirely clear to us what we ought to do, although we may lose sight of this when, for example, we are reluctant to fulfil a promise at our own expense. We may then turn to arguments to try to find evidence that we truly ought to fulfil our promise. According to Prichard, this demand for further evidence is illegitimate, as it is based on "the mistake of supposing the possibility of proving what can only be apprehended directly by an act of moral thinking."[22]

We can distinguish two related claims here about why moral philosophers would be mistaken to reflect on whether we ought to do what is moral. The first is that there is no need for substantial contemplation or argumentation to establish that we have a reason to be moral, as we already know this without all this effort. Prichard believes it is self-evident both that certain actions are right and that we ought to do these. According to him, we will know immediately whether an action is obligatory in a particular situation when we face that situation ourselves.[23] And where we remain unsure, this must be because we are not aware of all the relevant aspects of the situation – once we do have a complete impression of the action under consideration, we do not need to be shown a reason to do it.[24]

The second claim is that it is not merely unnecessary to look at the question through the lens of a philosopher, but that something goes *wrong* when we do this. The thought is that any argumentative support we can give to our natural belief that

[20] Prichard, 'Does Moral Philosophy Rest on a Mistake?', 29.
[21] Prichard, 36.
[22] Prichard, 36.
[23] Prichard, 37.
[24] Prichard, 28.

we ought to be moral would be of the wrong kind. Prichard claims that philosophers' answers to 'Why be moral?' fall into two categories: they either argue that we ought to perform moral actions because this leads to our own happiness, or they argue that we ought to perform moral actions because the action itself or its consequences are good.[25] He then tries to show that these answers are not adequate: the first kind of answer may motivate us to do what is moral but does not give us the right sense of obligation to do moral actions, while the second kind of answer does not focus on what we ought to do, but instead addresses the motive from which we should act.[26] In sum, Prichard judges it to be a mistake for moral philosophy to concentrate on 'Why be moral?' because (a) it is self-evident which actions morality requires and that we ought to perform these, and (b) any reason moral philosophers can give us to be moral, as the result of non-moral reasoning, must be of the wrong kind.

T. M. Scanlon identifies a similar yet subtly different pair of problems for 'Why be moral?'. Rather than argue that it is futile or inappropriate to try to answer this question at all, Scanlon brings up the worry that there is no satisfactory answer to it, as no reason we can provide for being moral is quite of the right sort. He sees it as problematic to answer that we ought to do what is morally required *just because* it is morally required. There is always a reason to do a morally right action, this reason being that it is right. We can interpret this to be the kind of answer that Prichard gives.[27] Yet, do we really get anywhere by showing that we have a moral reason to be moral? Scanlon does not think so: "this is surely not the kind of answer that is wanted: it simply takes the reason-giving force of moral considerations for granted."[28] It is trivially true that morality requires that we do what is morally required – the claim that we have a moral reason to be moral does not give us substantial information about what we ought to do.

However, Scanlon argues that the alternative strategy of establishing a non-moral reason to be moral does not fare any better, as this reason would be of the wrong kind.

---

[25] Prichard, 22.

[26] Prichard, 23–27.

[27] Scanlon has a different interpretation, on which Prichard avoids the problems by not giving any explanation of our reason to be moral (Thomas Scanlon, *What We Owe to Each Other* (Cambridge, Mass.: Belknap Press, 2000), 389, note 1). I take it that Prichard does provide an explanation, one that falls under Scanlon's category of giving a moral reason to be moral.

[28] Scanlon, 149.

For example, if we argue that it is ultimately in our own best interest to be moral, this reason for doing the morally right action has nothing to do with what makes this action morally right.[29] Consequently, we may think that a truly moral person would not do what she does for this particular reason. Scanlon thus identifies two pitfalls for responses to 'Why be moral': "a satisfactory answer to our question must not, on the one hand, merely say that the fact that an action is wrong is a reason not to do it; but it must, on the other hand, provide an account of the reason not to do it that we can see to be intimately connected with what it is to be wrong."[30] He calls this 'Prichard's dilemma'. Since I interpret Prichard to raise slightly different problems, I will refer to it as 'Scanlon's dilemma' instead.

It is important for Scanlon that the dilemma he describes does obtain, as it has a significant role in his argument for his own moral view. He argues that his contractualist view solves the dilemma and takes this victory to be substantial support for it. In Scanlon's version of contractualism, "an act is wrong if its performance under the circumstances would be disallowed by any set of principles for the general regulation of behavior that no one could reasonably reject as a basis for informed, unforced general agreement."[31] The reason to be moral he establishes is deeply connected to this account of wrongness: he argues that we have a reason to regulate our practical thought and behaviour in a way that others could not reasonably reject.[32] The reason we have not to break a promise, for example, is that it would be reasonable for others to object to this. This approach avoids the first horn of Scanlon's dilemma because it does not resort to the answer that we must do what is right because it is right – Scanlon presents a more substantial reason, which does not merely refer to the action's rightness. At the same time, his answer avoids the second horn of the dilemma in virtue of its close relation to the right-making feature of actions. In Scanlon's view, the reason we have to do morally required actions is ultimately also what makes these actions morally required.

I will not object here to Scanlon's claim that his view avoids the dilemma in this way; the part of his argument that I will now challenge, instead, is that there is a

[29] This assumes we do not accept ethical egoism.
[30] Scanlon, *What We Owe to Each Other*, 150.
[31] Scanlon, 153.
[32] Scanlon, 153–58.

dilemma to solve at all. I will take issue with the first horn: the problem that we cannot solve 'why be moral?' by showing that we have a moral reason to do what is morally right, as this is trivial.[33] The important step here is to distinguish between two kinds of normative reasons: formal reasons and authoritative reasons. We can highlight the difference between these by imagining norms that clearly do not give rise to authoritative reasons but nonetheless involve formal reasons. Imagine that I think of a set of rules that every person ought to live by, referring to these as the norms of schmorality. They are not meant to capture or replace the moral norms, but establish an additional type of requirement on people. Schmorality is comprised of a list of the most arbitrary and pointless rules I can think of: for example, one is required to scream 'Lilac!' at every third street one crosses, a scoop of strawberry ice cream may only be topped by a scoop of chocolate ice cream, and it is impermissible to live in a house with a flat roof. I will not connect my rules to awards or punishments and society in general will not enforce them by respecting compliers and shaming violators. All you will ever achieve by complying with these rules is that you have complied with the rules of schmorality.

Do we have a reason to do what schmorality requires? There is a sense in which we do. No matter how useless the rules of schmorality are, they do make certain courses of action correct or incorrect – (in)correct relative to schmorality, that is. And in virtue of making certain actions correct, the rules also give us a reason to perform these. You have a reason to scream 'Lilac!' during your third street crossing, the reason being that the rules of schmorality require this. This is an example of a formal reason. If there is a formal reason for you to φ, this merely means that a norm applies to you relative to which it is correct to φ. It is thus tautologically true that there is a formal reason to comply with a particular norm. If understood in the formal sense, 'You have a schmoral reason to do what schmorality requires' is obviously true, as it merely states

[33] To object to the second horn, one could argue that Scanlon is mistaken to assume a necessary connection between the normative reason we have to be moral and the appropriate motive for being moral. It may be appropriate to act on the reasons that explain why an action is morally required, rather than on the reasons that explain why we ought to do what is morally required. If saving the child from the pond is morally required because it prevents the child's death, that is also the right motivation to do it; this does not exclude the possibility that what gives an agent a reason to comply with this moral requirement is that it promotes their own preferences.

that it is correct relative to the rules of schmorality to do what is correct relative to the rules of schmorality.

Authoritative reasons are more significant than this. If a norm gives rise to authoritative reasons, it *matters* in some important sense – it is a norm that ought to be respected. Unlike formal reasons, authoritative reasons cannot legitimately be ignored. Authoritative normativity is both more exclusive and more forceful than formal normativity: not all formal reasons are authoritative reasons, and it is your authoritative reasons that you should pay attention to and act on. Considering that the rules of schmorality are entirely pointless, that compliance with them will not achieve anything of value, and that violations will have no negative effects for agents or others, it is very plausible that we have no authoritative reasons to comply with the rules of schmorality. If someone hears my list of rules and, sceptical that there is a point to following them, asks 'Why be schmoral?', I can give them a formal reason but not an authoritative one.[34]

We need to keep this distinction between formal and authoritative reasons in mind when we assess Scanlon's claim that it is trivial that we have a moral reason to be moral. If it is formal reasons we are thinking of, then 'we have a moral reason to do what is morally required' is tautological and does not tell us anything substantial about morality or what we ought to do. It is fair to call this result trivial. If it is authoritative reasons we are focused on, however, the answer that we have moral reasons do what is morally required is not trivial at all. It may be trivial that *if* there are authoritative moral reasons, they will be reasons in favour of doing what morality requires. Nonetheless, 'we have authoritative moral reasons to do what is morally required' is a meaningful claim, because it implies that moral reasons are indeed authoritative. This is not obviously the case. If moral reasons are authoritative, it matters what morality demands from us – whether an action is correct or incorrect relative to the norms of morality affects whether we really ought to do it. This view requires substantial assumptions about both morality and our authoritative reasons. Hence, if understood in the authoritative sense, an appeal to moral reasons to answer 'Why be moral?' is not

[34] I will expand on this sketch of the distinction between formal and authoritative normativity and defend the idea that formal normativity comes with reasons in the next chapter.

useless or uninformative. Therefore, the first horn of Scanlon's dilemma arises only if we are asking about formal reasons to be moral.

In contrast, 'Why be moral?' should be understood to ask specifically for authoritative reasons. Only this interpretation captures the meaningfulness and urgency of this question. Once it is clear that an action is morally required, the claim that we have a formal reason to do it adds nothing new. Establishing that there is an authoritative reason to do it, on the other hand, is informative and important. When someone who generally aims to live up to (what they take to be) morality's demands asks what reason they have to be moral, they are wondering if they are right to respect morality, or if they need to rethink which kinds of norms should inform their conduct. Their question is in the first place not about what is correct relative to some norm, but about which norms they should comply with – more specifically, whether the norms of morality are among them. The same applies when 'Why be moral?' is asked by an amoralist who does not currently pay heed to morality – they may acknowledge that certain actions are morally wrong, but are unconvinced that this gives them any significant reason not to do them. A successful answer to their question would show that they are making a serious mistake when they ignore moral requirements, because they do matter. In conclusion, since we can distinguish between formal and authoritative reasons, and because we should understand 'Why be moral?' to ask for the latter, it is not trivial to answer that we have a moral reason to be moral. With its first horn lost, Scanlon's dilemma does not obtain.

3. The self-evidence of reasons to be moral

With an awareness of the distinction between formal and authoritative reasons, as well as of the different types of internalism defined earlier, we can now respond to Prichard's complaint about 'Why be moral?'. Recall that, on my interpretation, his first objection to asking this question is that the answer is already obvious to us – since we already know what we ought to do in specific situations, or will know this once we find ourselves facing it, there is no point in philosophers investigating this question. This is slightly different from Scanlon's dilemma: while Scanlon raises a problem with

the appropriateness of possible answers to 'Why be moral?', Prichard focuses on knowledge, claiming that the answer is self-evident.

Perhaps Prichard sees the answer as self-evident for the same reason that Scanlon sees a certain answer to the question as trivial: because it is tautologically true that we have a moral reason to do what is moral. A tautology would certainly explain the self-evidence. However, here Prichard encounters the same objection I made to the first horn of Scanlon's dilemma: this tautology only arises if it is formal moral reasons we are thinking of. Since having a formal moral reason to φ just consists in there being a moral requirement on you to φ, it is self-evident that you have a formal moral reason to do what morality requires. No argumentation or philosophical reflection is needed to discover that you have a formal reason to do what a norm requires. If it were formal reasons to be moral we were looking for, we could agree with Prichard that there is no need for philosophers to spill any ink on this easily fulfilled quest. Things are not quite that simple, however: as I have argued, the real demand behind 'Why be moral?' is for authoritative reasons, not formal ones. To disregard this question by responding that it is self-evident that we have moral reasons to be moral, Prichard needs to prove that it is self-evident that moral reasons are authoritative. Below, I will explain that this is indeed Prichard's position and raise doubts about his argument for it. First, I will consider the broader idea that it is self-evident that we have an authoritative reason to be moral. For this, we should also keep the possibility in mind that we have an authoritative non-moral reason to be moral.

Is it self-evident that we have an authoritative reason to do what morality requires? To answer this, we need to get clearer on what it means for something to be self-evident, as well as on what can make this particular proposition self-evident. For these purposes, I will adopt the basic definition of self-evidence used by DePaul and Hicks: "A proposition P is self-evident just in case P is evident for any person S who understands P. (P is evident for S when P has an especially high degree of justification for S, where this is at least justification sufficient to know.)"[35] It should be noted that Prichard's standard seems to be even higher than this: he implies that the belief that we have a reason to do what is morally required has such a high degree of justification for us

[35] Michael DePaul and Amelia Hicks, 'A Priorism in Moral Epistemology', in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta (Stanford University, fall 2016), https://plato.stanford.edu/archives/fall2016/entries/moral-epistemology-a-priori/.

that there is simply no point in doubting it. There seem to be two ways to establish this: the first is to argue that we have intuitive knowledge of our authoritative reasons; the second is to argue that it is a conceptual truth that we have a reason to do what is morally required. Given the known problematic commitments involved in the theory that we can intuit normative facts, I will discuss only the second strategy.

There are two different paths to take to argue that it is a conceptual truth that we have a reason to do what is morally required: we can focus on moral concepts and try to uncover a commitment to authoritative normativity in them, or focus on authoritatively normative concepts to reveal that it is enclosed in their meaning that moral reasons are authoritatively normative. If the meaning of moral or normative terms necessitates that we have an authoritative reason to do what is morally required, then the answer to 'Do I really have a reason to do what is morally right?' is obvious to anyone who fully understands these terms. I will discuss these two options in turn.

To say that it is part of the meaning of moral terms that we have a reason to comply with moral requirements is to accept what I have introduced as conceptual internalism.[36] According to this view, 'you are morally required to φ' entails 'you have a reason to φ', just like 'it is morally impermissible to φ' entails 'you have a reason not to φ'. If conceptual internalism is true, normativity is simply built into moral concepts like these. Consequently, any norm we have no reason to comply with cannot be a moral norm. If it is formal reasons we are talking about, conceptual internalism is certainly true: it is indeed impossible that we have no formal reason to do an action which is morally required. The existence of a formal reason is contained in the concept of something being morally required. However, this is not the version of conceptual internalism we are interested in here: since our question is about the self-evidence of *authoritative* reasons to be moral, we need to examine whether conceptual internalism about authoritative reasons is true.[37]

Do our moral concepts come with built-in commitments to authoritative reasons, then? It is not clear that they do. The problem is not merely that there is no conclusive evidence that conceptual internalism is true; it is unclear what *could* even count as

[36] 'Conceptual reasons existence internalism' in full.

[37] There is thus another distinction to add to our classification of different internalisms: we can distinguish *conceptual formal reasons existence internalism* from *conceptual authoritative reasons existence internalism*.

conclusive evidence for it. Under the assumption that the meaning of moral terms is determined by ordinary speakers' use of them, a conceptual connection to authoritative normativity would have to be revealed by ordinary moral discourse. However, even if ordinary speakers do seem to connect claims about moral requirements to claims about reasons, this still need not show that conceptual internalism is true. There are several reasons for this. Firstly, it is important to distinguish conceptual internalism from what Goldsworthy calls *sociological internalism*, which is the thesis that it is generally believed to be true that morality has practical force.[38] Even if it was clear beyond any doubt that there is a universal belief that moral requirements are reason-giving, this would not establish that this is encoded in our moral concepts; the possibility remains that a common belief that we have a reason to keep our promises, for example, is explained by people's substantive view of reasons, rather than by what it means for keeping promises to be morally required. Therefore, *if* there is a prevalent belief that we have a reason to do what morality demands, this still leaves open whether conceptual internalism is true.

Secondly, the search for evidence for conceptual internalism is complicated by the distinction between formal and authoritative reasons. We are looking for evidence specifically for the claim that authoritative reasons are contained in our moral concepts. However, ordinary talking and thinking about morality and reasons does not get that specific. Normative terms like 'reason' and 'ought' are ambiguous: they are used to make claims about both authoritative and formal reasons, and it is normally not specified which is meant. This makes it less clear whether moral discourse involves a commitment to authoritative reasons. At least some moral claims merely seem to point out that there is a formal reason of morality to do something – for example, someone who says 'It is morally wrong to avoid tax, but I will do it anyway' could be taken to describe a moral norm without acknowledging an authoritative reason to comply with this norm. The issue is not just that users of normative language fail to clarify exactly what they have in mind; it is plausible that people do not have either an authoritative or a formal reading in mind when they say, for example, that we ought not lie. Furthermore, while ordinary patterns of behaviour suggest that people usually take moral requirements seriously, it would be a step too far to conclude from this

[38] Goldsworthy, 'Externalism, Internalism and Moral Scepticism', 41.

that authoritative reason-givingness must be conceptually linked to morality. As I will discuss in later chapters, there are alternative ways to make sense of the importance we ascribe to morality. In the end, it is far from clear that authoritative normativity, and not just formal normativity, is built into our moral concepts.

A different problem is that the truth of conceptual internalism itself cannot establish that we do have a reason to keep our promises, save children from ponds, refrain from stealing, or do any other things usually considered to be morally required. This is because conceptual internalism only entails that *if* keeping promises, saving children, et cetera is morally required, *then* we necessarily have a reason to do this. Clearly, this is only half the story: we still need to establish that these actions, or anything at all for that matter, are indeed morally required. Conceptual internalism is a thesis about what normative power moral demands must have over us, but it does not tell us whether moral demands actually exist. As a matter of fact, conceptual internalism has been used to argue for the opposite – in chapter five, I will discuss how some moral error theorists use the conceptual claim that we necessarily have categorically authoritative reasons to do what is morally required to conclude that nothing is morally required.

Consider again Prichard's example of a person who, faced with an action that goes against her own interest, wonders whether she has been under an illusion in thinking she ought to do it. Now imagine that conceptual internalism is true, and our reluctant agent knows this. She thus knows that if she is under a moral obligation to do the action, she has a reason to do it. However, rather than making her question evaporate, this just shifts its focus: her aim is now to discover whether there is indeed a moral obligation to do the action under review. In this way, even if conceptual internalism is true, it does not provide an answer to whether we have a reason to do what we take to be morally right – it merely switches the question to one about whether these actions actually are morally right. If the answer to this is not self-evident, then it is not self-evident that we have a reason to keep promises and save drowning children, whether normativity is built into our moral concepts or not. In absence of a story of how it could be self-evident that some actions are indeed morally required, scepticism about us having a reason to do these actions still stands.

An alternative way in which it could be conceptually true that we have a reason to be moral is if it is contained in the concept of authoritative normativity that moral

reasons are authoritative. We can interpret Prichard to make this claim. To explain why people continue to ask 'Why ought I to do what is my duty?' despite this question being absurd, Prichard points out that there are two very different senses of 'ought': a moral sense and a non-moral sense. In his view, the non-moral 'ought' refers to what is instrumental in achieving a certain aim an agent has: used like this, 'ought to be done' means "proper for realizing the agent's purpose".[39] In contrast, when 'ought' means 'morally ought', it is not related to agents' purposes. When, faced with a morally required action we are reluctant to do, we ask whether we ought to be moral, we are using the non-moral, instrumental 'ought'. On one plausible reading of Prichard, the problem he sees with this is that the non-moral 'ought' is not normative.[40] Given that non-moral ought-claims merely describe means-end relations, they are descriptive rather than prescriptive.[41] To say that a poisoner, in order to be successful, ought to use a larger dose[42] is not to say that the poisoner is under some kind of obligation to use a larger dose, or that he does anything wrong if he does not do this. In my terms, Prichard believes that non-moral reasons are merely formal and not authoritative.

The next step is to establish that moral reasons, in contrast, are authoritative. On Thomas Hurka's interpretation, Prichard is a conceptual minimalist who only accepts one kind of genuine (i.e. authoritative) 'ought'. As explained, claims with the instrumental 'ought' can be reduced to claims about what means will lead to a certain end. What is left as authoritatively normative is a categorical 'ought', which is not conditional upon agents' ends. Thus, according to Hurka, Prichard's position is that "The only genuine 'oughts' are categorical, and since (following Kant) all categorical 'oughts' are moral, there are only moral 'oughts' or moral requirements."[43] It is presented as contained in the concept of normativity that moral oughts and reasons are genuinely normative, and non-moral oughts and reasons are not. It follows from this

[39] Prichard, 'Manuscript on Morals', 126.

[40] This reading is questioned in Evan Tiffany and Sam Black, 'Moral Philosophy Does Not Rest on a Mistake: Reasons to Be Moral Revisited', in *Reasons to Be Moral Revisited*, ed. Evan Tiffany and Sam Black, Canadian Journal of Philosophy Supplementary, v. 33 (Calgary: University of Calgary Press, 2007), vii–xl.

[41] Thomas Hurka, 'Underivative Duty: Prichard on Moral Obligation', *Social Philosophy and Policy* 27, no. 02 (2010): 111–34.

[42] As Prichard does in Prichard, 'Manuscript on Morals', 126.

[43] Hurka, 'Underivative Duty', 120.

that it is a mistake to ask what non-moral reason we have to do what morality requires: to give such a reason is not to show that we have an authoritative reason to be moral. Asking for moral reasons to be moral, on the other hand, is also pointless: moral reasons are by definition reasons in favour of being moral.[44] If it is self-evident, in virtue of being conceptually true, that moral reasons are authoritative and non-moral reasons are not, then there is no answer to find to 'Why be moral?' that is both valid and useful.

This argument relies on an overly simplistic view of types of reasons. There is an endless number of systems of norms. All these norms can come with formal reasons, which may or may not also be authoritative. Consequently, it is not true that the only non-moral requirements are the requirements following from the means-end principle. Apart from instrumental reasons, the list of non-moral reasons also includes epistemic reasons, aesthetic reasons, reasons of rationality, etiquette reasons, football reasons, et cetera. These may be merely formal reasons, but Prichard does not give us a reason to think that they are not authoritative, since he only considers instrumental reasons to be an alternative to moral reasons.

Furthermore, many of these additional types of reasons are categorical. Because it does not depend on my desires whether the norms of etiquette apply to me, I have reasons of etiquette to perform certain actions regardless of my desires.[45] This means that even if we were to agree that only categorical reasons can be authoritative, this does not warrant the conclusion that moral reasons and *only* moral reasons are authoritative, as there are still different options left. Hurka claims that it is unobjectionable to simply call any genuinely normative categorical 'ought' a moral 'ought', because nothing substantive can turn on this terminology.[46] I believe this is false: it implies that if there are any authoritatively normative categorical requirements, moral requirements are authoritatively normative. However, if we only have authoritative categorical reasons of a different kind, such as rational, epistemic or aesthetic ones, it may not be the case that we have authoritative reasons to comply with moral requirements. Therefore, it does make a substantial difference whether our authoritative oughts and

[44] On this view, the first horn of Scanlon's dilemma does obtain even if we are talking about authoritative rather than formal reasons, because it is assumed not to be an open question whether moral reasons are authoritative.

[45] See Philippa Foot, 'Morality as a System of Hypothetical Imperatives', *The Philosophical Review* 81, no. 3 (1972): 305–16.

[46] Hurka, 'Underivative Duty', 121.

reasons are moral ones or not, and we cannot simply call all authoritative categorical reasons 'moral' without risking unwarranted implications. With a less simplistic picture of the normative landscape in mind, the argument Hurka takes Prichard to make for the normativity of moral reasons fails.

A last problem for this strategy to show that it is beyond question that we have a reason to be moral is that we can seriously doubt that it is obvious to everyone that moral reasons are authoritative and instrumental reasons are not. Most of us take our instrumental reasons very seriously. Considerations about what the means to our ends are take a central role in practical deliberation. In the cases where we do give priority to moral considerations, the underlying assumption seems to be that moral reasons override instrumental ones, not that it does not speak in favour of an action at all that it is instrumental to achieving our ends. Considering the significant role instrumental reasons have in our deliberation, motivation, and normative discourse, it is very implausible that it is self-evident that instrumental reasons are not authoritatively normative. It is important to remember that it is not sufficient for Prichard if, despite ordinary beliefs, it is true after all that moral reasons are authoritative and all other reasons are not; only if this is self-evident will Prichard's complaint about 'Why be moral?' be warranted. For the reasons given, I strongly doubt that it is.

4. Conclusion

I have explored which question Prichard considers to be the mistaken question of moral philosophy, what we should understand this question to ask for, and how it relates to a question about the content of morality. I targeted Scanlon's dilemma for providing a reason to be moral, arguing that when we are focused on authoritative reasons, as we should be, it is not trivial to say that we have moral reasons to be moral. Prichard's own view on the problem with 'Why be moral?' does not fare better: I have found no successful account of how the answer could be so self-evident that it is a mistake to ask this question at all. It is not clear that it is contained within either our moral concepts or the concept of authoritative normativity that we have an authoritative reason to do what is morally required. Instead, there is plenty of work to do for philosophers here. 'Why be moral?' is a question about our authoritative reasons,

taking us to a complex debate which does not have straightforward solutions that are perfectly obvious to anyone who considers the issue. There is a point to investigating the question because it involves achieving a better understanding of what the difference is between formal and authoritative normativity, which reasons are authoritative and why, whether the reasons we have to be moral are truly authoritative, and what it means if they are not. Moral philosophy, and the debate on moral normativity in particular, does not rest on a mistake.

CHAPTER 2

Demystifying Authoritative Normativity

0. Introduction

In the previous chapter, I touched upon the difference between formal and authoritative normativity and concluded that answering 'Why be moral?' requires an understanding of this distinction. Yet, its importance goes beyond the subject of moral normativity: in recent years, it has become clear that when we ask about the normativity of the law, meaning, logic, rationality, and other domains, it is specifically authoritative normativity we are interested in. In spite of the growing popularity of this concept in the metanormative literature, it remains clouded in mystery. Instead of clearly delineating this form of normativity, it is usually merely described with rough characterisations and metaphors meant to convey a sense of the contrast with formal normativity. This constitutes a precarious starting point for metanormative debates. Two problems with current discussions of authoritative normativity stand out.

First, it is often conflated with the concept of ought *simpliciter*. Because these have not been separated in the literature, it is common for the focus of discussions on authoritative normativity to switch unintentionally between these two normative

concepts. Once we recognise that these are not the same, it becomes clear that the debate on this topic has been unhelpfully clouded so far.

Second, the debate is shaped by unwarranted preconceptions about the nature of normative authority. Despite our limited grasp of the concept, it has been widely assumed that authoritative normativity is by definition irreducible. As a result, metaphysically naturalist explanations of authority are deemed to be out of the question. This not only skews the debate in favour of non-naturalist theories, but also plays into the hands of normative error theorists: when authority is conceived of as a rather extraordinary property, it becomes harder to establish that it is instantiated. In order to have a fair debate and avoid undue scepticism, we need to be more critical about what belongs to the bare concept of authoritative normativity, and what is a substantive conception of it that calls for a defence.

The aim of this chapter is to demystify authoritative normativity in the following ways. First, I will introduce the basic idea of the distinction between formal and authoritative normativity by contrasting it with the view that there is only one kind of normativity. Second, I will provide an overview of common conceptions of the defining feature of authoritative normativity and argue that these fail. Third, I will show that the concept under discussion here is not the same as that of the ought *simpliciter*. Fourth, I will turn to the assumption, central to Jonas Olson's normative error theory and shared by many, that authoritative normativity is defined by irreducibility. I will distinguish two versions of this assumption and argue against both. With these demanding preconceptions out of the way, there is conceptual room for metaphysically naturalist theories that take the mystery out of authoritative normativity.

1. Two kinds of normativity

Norms are ubiquitous. By 'norms' I mean standards, rules, principles or conventions that involve correctness conditions, rather than statistical normalities. Norms involve correctness conditions because they categorise our behaviour and attitudes as correct or incorrect, depending on whether they are in accordance with the norm in question. Wearing socks in sandals is incorrect, relative to fashion, just in virtue of it violating

the fashion norm not to wear socks in sandals.[1] We can use the term 'system of norms' for a set of norms that belong together in a single domain, such as morality, the law, logic, and chess. While the content of some norms – such as those of prudence and logic – does not seem to be up to us, other norms are simply the result of conventions, bureaucratic procedures, or decisions by the Rules Committee of an ancient golf club. That norms have such a variety of sources gives some reason to expect that they do not all have the same normative status. Some norms appear to be binding for us in a way that others are not. This is not a matter of motivational power: if I am disposed to follow the fashion norms but refuse to take the requirements of prudence seriously, it still seems true that prudence affects what I really should do, while the norms of fashion may not.

There are two different ways to conceptualise this distinction between norms that do and norms that do not have a significant normative force. The first option is to say that not all norms are normative. On this view, the fact that a norm N requires action φ, thereby making φ-ing correct relative to N, does not entail that N is normative; it merely means that this norm appears to be normative, which it may in reality not be. If the requirements of prudence matter for what we ought to do in a way that fashion norms do not, this is because the norms of prudence are genuinely normative and fashion norms are not. There is only one kind of normativity, involving one kind of ought and reasons, and it does not apply to all norms (or, possibly, to any).

Within this picture, different interpretations of normative claims relating to non-normative norms – e.g. 'you ought not wear socks in sandals' – are possible. One straightforward interpretation is that these are false. If we suppose that there is strictly one kind of normativity, then all normative claims must refer to this. It is then a mistake to talk of oughts or reasons produced by a non-normative norm. However, Daniel Wodak has argued for a different interpretation: instead of expressing a false belief, our normative claims about (some or all) non-normative norms express a fictionalist attitude.[2] In this view, fashion norms are not truly reason-giving. However, we can see the fashion norms as giving rise to a fiction in which fashion reasons do exist. We can

[1] Uncontroversial and stable norms of fashion are surprisingly difficult to find. Readers who do not recognise this particular norm are free to keep a different example in mind.

[2] Daniel Wodak, 'Mere Formalities: Fictional Normativity and Normative Authority', *Canadian Journal of Philosophy* 49, no. 6 (2019): 828–50.

step into this fiction and speak about what we ought to do from its perspective without genuinely taking it to be true that we ought not wear socks in sandals. We can thus take a fictionalist attitude towards non-normative norms and accept, without believing, normative propositions about them.

The problem with denying that all norms are normative is that this obscures what they do have in common: all norms have an action-guiding nature. As David Copp points out, that morality "prescribes, recommends, and evaluates actions and choices" is a feature it shares with etiquette and aesthetic norms.[3] The point of norms is not to describe, but to determine which actions or attitudes are correct, right, legitimate, or ought to be chosen. This applies no less to norms whose classification of something as correct or incorrect seems irrelevant. Consider Derek Baker's response to the hypothesis that etiquette is not normative: "But in what sense are reasons of etiquette not normative? They determine ought-facts that are prescriptive, not predictive. They favor and disfavor options. Claims about what you have reason of etiquette to do are not falsified by non-compliance. Non-compliance licenses criticism."[4] There is a natural way of capturing this set of features norms have: norms are normative. It is their normativity that unites norms and contrasts them with anything else. Nicholas Southwood illustrates this for the norms of rationality:[5]

> There is at least one way in which rationality is normative, namely, that it is constituted by requirements. Rationality requires things of one, and we can violate these requirements (...). This is a way that rationality is normative that, say, the history of the printing press, the content of a Woolworths supermarket, and Bach's Brandenburg Concertos aren't.[6]

We can use a sense of 'normative' in which it merely involves there being a standard of correctness.[7] In this sense, all norms are necessarily normative. Normative terms like

---

[3] David Copp, 'Moral Naturalism and Three Grades of Normativity', in *Morality in a Natural World: Selected Essays in Metaethics* (Cambridge: Cambridge University Press, 2007), 14.
[4] Derek Baker, 'Skepticism About Ought Simpliciter', in *Oxford Studies in Metaethics 13*, ed. Russ Shafer-Landau (Oxford: Oxford University Press, 2018), 239.
[5] This assumes a conception of rationality as a system of norms, which I discuss below.
[6] Nicholas Southwood, 'Vindicating the Normativity of Rationality', *Ethics* 119, no. 1 (2008): 11.
[7] John Broome, 'Is Rationality Normative?', *Disputatio* 2, no. 23 (2007): 162.

'ought' also have a sense that merely invokes a standard of correctness and is therefore appropriate to use for any norm. In one sense, it is true that I ought not wear socks in sandals – this merely means that a norm classifies my wearing socks in sandals as incorrect. However, this is not to say that something being incorrect relative to fashion has the same significance as it being incorrect relative to morality; we can still conceive of different norms as having a different normative status.

This is accommodated by the second way of conceptualising the difference in normative force between norms: all norms are normative, but not all norms are normative in an especially significant way. We can distinguish two different kinds of normativity: formal and authoritative normativity. Formal normativity merely involves standards of correctness. All norms are formally normative, in virtue of requiring us to perform certain actions or attitudes and not others. Authoritative normativity is the more significant normative power that norms can have on top of this. This distinction has recently become commonly recognised in metanormative debates, from metaethics and jurisprudence to the normativity of meaning. What I call 'formal normativity'[8] has also been discussed as norm-relativity[9] and generic[10] or rule-involving[11]

[8] After Tristram McPherson, 'Against Quietist Normative Realism', *Philosophical Studies* 154, no. 2 (2011): 223–40; David Plunkett and Scott Shapiro, 'Law, Morality, and Everything Else: General Jurisprudence as a Branch of Metanormative Inquiry', *Ethics* 128, no. 1 (2017): 37–68; Derek Baker, 'The Varieties of Normativity', in *The Routledge Handbook of Metaethics*, ed. Tristram McPherson and David Plunkett (New York: Routledge, 2017), 567–81; Preston J. Werner, 'Why Conceptual Competence Won't Help the Non-Naturalist Epistemologist', *Canadian Journal of Philosophy* 48, no. 3–4 (2018): 616–37; Jack Woods, 'The Authority of Formality', in *Oxford Studies in Metaethics 13*, ed. Russ Shafer-Landau (Oxford: Oxford University Press, 2018), 207–29; David Enoch, 'Is General Jurisprudence Interesting?', in *Dimensions of Normativity: New Essays on Metaethics and Jurisprudence*, ed. David Plunkett, Scott J. Shapiro, and Kevin Toh (Oxford: Oxford University Press, 2019), 65–86; David Plunkett, 'Robust Normativity, Morality, and Legal Positivism', in *Dimensions of Normativity: New Essays on Metaethics and Jurisprudence*, ed. David Plunkett, Scott J. Shapiro, and Kevin Toh (Oxford: Oxford University Press, 2019), 105–36; George Letsas, 'How to Argue for Law's Full-Blooded Normativity', in *Dimensions of Normativity: New Essays on Metaethics and Jurisprudence*, ed. David Plunkett, Scott J. Shapiro, and Kevin Toh (Oxford: Oxford University Press, 2019), 165–86.

[9] Anandi Hattiangadi, *Oughts and Thoughts* (Oxford: Oxford University Press, 2007), 37.

[10] Copp, 'Moral Naturalism and Three Grades of Normativity'.

[11] Derek Parfit, *On What Matters: Volume One* (Oxford: Oxford University Press, 2011), 144–45.

normativity. Authoritative normativity[12] is also known as robust,[13] full-blooded,[14] substantive,[15] true[16] or reason-involving[17] normativity. It has proven difficult to express in a precise way what this distinction comes down to. We are left with rather open-ended characterisations of the difference in normative force. Like David Enoch, we may understand authoritative norms to be norms we are bound by, in some way to be specified: "Morality is not just a game that generates criteria of correctness – it is *the right* game, a game whose correctness conditions you can't escape by refusing to play it."[18] Richard Joyce speaks of "practical oomph".[19] Following Baker, we could focus on a contrast in the legitimacy of violating norms: "A person can ignore or even wilfully violate [merely formal norms] (...) without any implication that she is guilty of some sort of mistake or that her behavior is in any interesting sense defective."[20] In the end, the basic idea is that the requirements of norms with authoritative normativity are significant for what we ultimately ought to do in a way that the requirements of other norms are not. Norms tell us what to do, but only authoritative norms "*really*" determine what we ought to do.[21] In Linda Radzik's words, such a norm "legitimately restricts us" and "makes a genuine claim on us".[22] What we are interested in is a distinction, as Stan Husi puts it, "between the formal feature of being directive in

[12] After Copp, 'Moral Naturalism and Three Grades of Normativity'; Baker, 'The Varieties of Normativity'; Barry Maguire and Jack Woods, 'The Game of Belief', *The Philosophical Review* 129, no. 2 (2020): 211–49.

[13] McPherson, 'Against Quietist Normative Realism'; Plunkett and Shapiro, 'Law, Morality, and Everything Else'; Werner, 'Why Conceptual Competence Won't Help the Non-Naturalist Epistemologist'; Plunkett, 'Robust Normativity, Morality, and Legal Positivism'.

[14] Enoch, 'Is General Jurisprudence Interesting?'; Letsas, 'How to Argue for Law's Full-Blooded Normativity'.

[15] Woods, 'The Authority of Formality'; Susanne Mantel, 'Do Epistemic Reasons Bear on the Ought Simpliciter?', *Philosophical Issues* 29, no. 1 (2019): 214–27.

[16] John Broome, *Rationality through Reasoning* (Oxford: Wiley-Blackwell, 2013), 11.

[17] Parfit, *On What Matters*, 144–45.

[18] David Enoch, 'Non-Naturalistic Realism in Metaethics', in *The Routledge Handbook of Metaethics*, ed. Tristram McPherson and David Plunkett (New York: Routledge, 2017), 33.

[19] Richard Joyce, *The Evolution of Morality* (Cambridge, Mass.: The MIT Press, 2007), 62.

[20] Baker, 'The Varieties of Normativity', 568.

[21] Baker, 568–69.

[22] Linda Radzik, 'A Coherentist Theory of Normative Authority', *The Journal of Ethics* 6, no. 1 (2002): 21.

character and the substantive feature of directing with authority."[23] Unfortunately, as I will now show, the different more specific diagnoses that have been offered of this distinction are either inaccurate or insufficient.

2. Defining authoritative normativity

The *locus classicus* of the distinction between formal and authoritative normativity is arguably Philippa Foot's 'Morality as a System of Hypothetical Imperatives'. Foot points out that even relatively insignificant norms apply categorically to us: the rules of etiquette do not fail to apply to someone who does not care about what she should do from the point of view of etiquette. Yet, these norms do not have the "inescapability" or "special dignity and necessity" that is typically ascribed to morality; therefore, there is supposed to be a normative difference between etiquette and morality that goes beyond the difference between hypothetical and categorical imperatives.[24] In this way, Foot has shown that authoritative normativity cannot be identified as normativity that is categorical.

Foot ties the property of being reason-giving to authoritative normativity only: while she describes etiquette as being normative and giving rise to a 'should', she does not take considerations of etiquette to be reasons for action.[25] This conception of authority as reason-givingness is common. For example, Broome defines the authoritative sense of normativity as follows: "a requirement on you to *F* is normative if and only if it constitutes a reason for you to *F*".[26] Derek Parfit also sees reasons as exclusive to the authoritative side, as he portrays the distinction in terms of a rule-involving and reason-involving conception of normativity:

[23] Stan Husi, 'Why Reasons Skepticism Is Not Self-Defeating', *European Journal of Philosophy* 21, no. 3 (2013): 425.
[24] Foot, 'Morality as a System of Hypothetical Imperatives', 1972, 308.
[25] Foot, 309.
[26] John Broome, 'Requirements', in *Homage à Wlodek: Philosophical Papers Dedicated to Wlodek Rabinowicz*, ed. Toni Rønnow-Rasmussen et al., 2007, 3, https://www.fil.lu.se/hommageawlodek.

> On the *reason-involving* conception, normativity involves reasons or apparent reasons.[27] On the *rule-involving* conception, normativity involves requirements, or rules, that distinguish between what is *correct* and *incorrect*, or what is *allowed* and *disallowed*.[28]
>
> These conceptions of normativity are very different. On the rule-involving conception, we can create new normative truths merely by introducing, or getting some people to accept, some rule. Legislators can create laws, and anyone can create the rules that define some new game. (...) In contrast, on the reason-involving conception, there is normativity only when there are normative reasons or apparent reasons. We cannot create such reasons merely by getting people to accept some rule.[29]

Parfit treats reasons as a neutral criterion which can be used to compare the normative power of different requirements. This enables him to point to reasons as the fundamental source of authority: in his view, a norm matters if and only if we have a reason to care about it.[30]

However, formal normativity does seem to involve reasons too. Just like there is a formal sense of 'ought', there is a merely formal sense of 'reason'.[31] All norms can involve a web of requirements, oughts and reasons, regardless of whether they are authoritatively or merely formally normative. When the rules of chess give me two options for my next move, I may have a chess reason to advance a pawn and a stronger chess reason to move my king to a safe place. I can alert you to a fashion norm by saying 'You ought not wear socks in sandals', as well as by saying 'You have a reason not to wear socks in sandals'. Although the second claim is less natural, it does not add

[27] In Parfit's framework, merely apparent reasons are reasons provided by false beliefs. Parfit, *On What Matters*, 35.
[28] Parfit, 144.
[29] Parfit, 145.
[30] Parfit, 146–48.
[31] Stephen Finlay, 'Defining Normativity', in *Dimensions of Normativity: New Essays on Metaethics and Jurisprudence*, ed. David Plunkett, Scott J. Shapiro, and Kevin Toh (Oxford: Oxford University Press, 2019), 205.

anything new, and does not imply that the fashion norm is authoritative.[32] Understood in this way, reasons are not a special, necessarily authoritative normative entity; we can bring them into existence just by creating rules. Norms can come with their own internally recognised set of reasons. This does not settle whether we *really* ought to do what some norm requires of us.[33]

The claim that only authoritative norms produce reasons seems to rely on the view that there is only one kind of reasons: reasons *simpliciter*.[34] In contrast, we can recognise a range of different kinds of reasons, all connected to a particular normative system: there are moral reasons, prudential reasons, epistemic reasons, football reasons, fashion reasons, and so on. We can call these subscripted reasons or reasons-of-a-kind.[35] Oughts can likewise be subscripted: when the norms of etiquette require me to φ, I $\text{ought}_{\text{etiquette}}$ to φ, even if I simultaneously $\text{ought}_{\text{moral}}$ not to φ. Distinguishing different kinds of reasons in this way allows us to ascribe special normative power to some reasons and not others. We can conceive of moral reasons as inescapably binding while at the same time describing fashion reasons as normatively insignificant. The crucial question is not which norms provide reasons, but which kinds of reasons are authoritative. Therefore, we cannot define authoritative normativity as the form of normativity that involves reasons.

Another hypothesis is that authoritative normativity is defined by prescriptivity. In debates on whether meaning is normative, it is commonly suggested that the kind of normativity of interest, which goes beyond there being correctness conditions for the use of a term, is prescriptive in nature.[36] But what does it mean to be prescriptive?

---

[32] Richard Joyce, *The Myth of Morality* (Cambridge: Cambridge University Press, 2001), 40.

[33] Stephen Darwall, 'Reasons, Motives, and the Demands of Morality: An Introduction', in *Moral Discourse and Practice: Some Philosophical Approaches*, ed. Stephen Darwall, Allan Gibbard, and Peter Railton (Oxford: Oxford University Press, 1997), 306; Husi, 'Why Reasons Skepticism Is Not Self-Defeating', 426.

[34] I discuss the relation between authoritative normativity and the ought *simpliciter* below.

[35] The term 'reasons-of-a-kind' is used in Copp, 'Moral Naturalism and Three Grades of Normativity', 31.

[36] E.g. Kathrin Glüer, 'Sense and Prescriptivity', *Acta Analytica* 14, no. 23 (1999): 111–28; Kathrin Glüer and Åsa Wikforss, 'Against Content Normativity', *Mind* 118, no. 469 (2009): 32; Anandi Hattiangadi, 'Is Meaning Normative?', *Mind & Language* 21, no. 2 (2006): 222–24; Hattiangadi, *Oughts and Thoughts*, 37; Jeffrey Kaplan, 'The Problem with Descriptive Correctness', *Ratio* 33, no. 2 (2020): 79–86.

'Prescriptive' and its counterpart 'descriptive' seem in the first place to be labels of two linguistic categories of statements: descriptive statements describe something, while prescriptive statements dictate or recommend something. Imperatives are paradigmatically prescriptive. Yet, this form of prescriptivity is not exclusive to claims concerning authoritative normativity; 'Do not eat peas with a spoon!' is just as much an imperative as 'Do not steal!' is. In general non-philosophical usage, the main definition of 'prescriptive' is simply 'serving to prescribe',[37] where 'to prescribe' means 'to lay down a rule'.[38] Since this is precisely the point of formally normative claims, this does not make being prescriptive a defining feature of authoritatively normative claims. 'Prescriptive' could have a different usage within philosophy, but if so, it is far from clear what its exact meaning is here.[39] It can be difficult to separate descriptive from prescriptive judgements, and we may draw the lines in different places. On the one hand, we can point out that 'You ought to eat peas with a fork' is ultimately nothing more than a description of what is correct relative to some norm. On the other hand, the claim does revolve around an 'ought', which is a classic feature of prescriptions. Claims concerning formal normativity can have all the signs of a prescriptive claim; to say that they are nonetheless merely descriptive is to use a particular understanding of 'prescriptive' that calls for further explication. For these reasons, it is not helpful, and seemingly even inaccurate, to define authoritative normativity as the form of normativity that is prescriptive in nature.

Finally, some treat a certain connection to rationality as the defining feature of authoritative normativity. Copp formulates two such proposals for a definition of authoritative normativity.

> *Authoritative reasons proposal*: "requirements of some kind K have authoritative normativity just in case, necessarily, there is an authoritative reason to act as one is K-required to act – a reason that any rational

[37] *Merriam-Webster.com Dictionary*, s.v. "prescriptive," accessed January 7, 2021, https://www.merriam-webster.com/dictionary/prescriptive.

[38] *Merriam-Webster.com Dictionary*, s.v. "prescribe," accessed January 7, 2021, https://www.merriam-webster.com/dictionary/prescribe.

[39] Interestingly, the Oxford Dictionary of Philosophy does not have an entry on 'prescriptive' or 'prescriptivity'.

> person would take into account, if she were aware of it, just in virtue of being rational."[40]
>
> *Closed question proposal*: "requirements of kind K have authoritative normativity just in case, necessarily, no fully rational person who is thinking clearly can take there to be any serious practical question of whether to act in accord with K requirements."[41]

I will raise two problems with defining authoritative normativity in terms of rationality, which are tied to two common but distinct conceptions of rationality.

First, being rational is sometimes understood to be a matter of responding appropriately to one's reasons, whatever their source. We must understand 'reasons' to refer to authoritative reasons here. A rational person would take into consideration, be motivated by, or act on her authoritative reasons (if she is or could be aware of them).[42] This idea involves many contested details that I will not address here. The important point is this: on this conception, rationality is defined in terms of authoritative reasons. Therefore, it will not do to define authoritative normativity in terms of rationality; such a definition is circular and will not get us anywhere. To say that it is irrational to ignore authoritative reasons is simply to reformulate the familiar point that authoritative reasons have a normative power that merely formal reasons lack, without explaining what this difference between authoritative and formal reasons consists in.

Second, some conceive of rationality as its own system of norms. Just like morality and etiquette, rationality consists of a set of norms that require actions or attitudes from us. While there are different views on what these norms are, the general consensus is that they are focused on creating coherence among one's attitudes.[43] For example, one norm of rationality may be that if you believe P, and you believe that if P then Q, then you ought to believe Q. Rationality also seems to require that if you intend to φ, and you believe that φ-ing requires ψ-ing, then you intend to ψ. On this picture,

---

[40] Copp, 'Moral Naturalism and Three Grades of Normativity', 29.

[41] Copp, 29–30.

[42] Being rational is sometimes defined as responding appropriately to what one has most reason, all things considered, to do. This conception relates rationality to the ought *simpliciter*, discussed below.

[43] Southwood, 'Vindicating the Normativity of Rationality', 10.

being rational specifically requires responding appropriately to one's rationality reasons – the reasons provided by the norms of rationality. However, if rationality is simply a system of norms, it is an open question whether it is normative – that is, normative in the authoritative sense.[44] Copp points out that the fact that a rational person would take rationality reasons seriously does not entail that they are authoritative; notice that the fact that a moral person would take moral reasons seriously also does not settle whether morality is authoritative.[45] It is possible that rationality is indeed authoritative, but it is as least conceivable that its norms are merely formal. Therefore, on this conception of rationality, it is not accurate to define authoritative normativity in terms of rationality. In conclusion, it is either unhelpful or incorrect to define authoritative normativity – in contrast to formal normativity – as categorical, reason-giving, prescriptive, or as the kind of normativity that is irrational to ignore.

3. Ought *simpliciter*

An obstacle standing in the way of a clear understanding of authoritative normativity is that this term is also associated with a subtly different concept, which has not been explicitly teased apart from the formal/authoritative distinction I am concerned with. We can find this alternative normative concept in Tristram McPherson's discussion of authoritatively normative concepts. McPherson is not concerned with the authority of particular norms, requirements or reasons, but asks instead why certain normative *concepts* have authority – in particular, the concept he calls 'practical ought'. Its more common name is 'ought *simpliciter*'.

McPherson illustrates the role of this concept with a scenario of a classic deliberative conflict:

> *Sticky Situation* You find yourself in a sticky situation. You conclude that morality requires you to stay and help, while prudence dictates that

[44] John Broome has been influential in raising this question, e.g. in Broome, 'Is Rationality Normative?' See also Niko Kolodny, 'Why Be Rational?', *Mind* 114, no. 455 (2005): 509–63; Southwood, 'Vindicating the Normativity of Rationality'.
[45] Copp, 'Moral Naturalism and Three Grades of Normativity', 32.

> you take the money and run. Torn, you ask yourself: *given all of this, what ought I to do?*[46]

In Sticky Situation, it does not seem appropriate to answer the question by using an ought subscripted to a certain domain of norms. It is obvious that you $\text{ought}_{\text{prudential}}$ to take the money and run, while you $\text{ought}_{\text{moral}}$ to stay and help. What we appear to need here is an unsubscripted ought that tells us what is to be done overall, outside of the perspectives of particular norms. When systems of norms make conflicting demands, this ought *simpliciter* settles what you really ought to do. The ought *simpliciter* is supposed to have an authority over and above that of subscripted oughts: "[in Sticky Situation,] it seems most plausible to read 'ought' as expressing a concept that purports to wear a distinctive normative authority on its sleeve, in a way that even moral and prudential 'ought's do not."[47] This is the type of authority McPherson is interested in.

McPherson's account of the concept of ought *simpliciter* focuses on its relation to the activity of non-arbitrary selection. In cases of conflict between systems of norms, we need a way to choose between them that is not based on irrelevant considerations or otherwise merely random. McPherson leaves open what choosing non-arbitrarily amounts to and which considerations count as relevant. His claim is that an agent ought *simpliciter* to φ if and only if a successful instance of the activity of non-arbitrary selection – whatever that is like – results in selecting the option to φ.[48] McPherson then argues that the authority of the ought *simpliciter* is explained by the role it plays in the activity of non-arbitrarily selecting what to do: "it is the concept of a norm that is *the norm to appeal to* in the context of non-arbitrary selection. The fact that the concept PRACTICAL OUGHT satisfies this job description is what *constitutes* its distinctive authoritativeness."[49]

Although McPherson uses the term 'authoritative normativity' to point to the special force of the ought *simpliciter*, the concept he analyses is different from the

[46] Tristram McPherson, 'Authoritatively Normative Concepts', in *Oxford Studies in Metaethics 13*, ed. Russ Shafer-Landau (Oxford: Oxford University Press, 2018), 254.
[47] McPherson, 255.
[48] McPherson, 266.
[49] McPherson, 267.

concept of authoritative normativity I am concerned with.[50] The formal/authoritative distinction can be used to compare the normative power of norms, but does not necessarily settle which norm we should follow in case of a conflict. If morality and prudence are both authoritative rather than merely formal, then the concept of authoritative normativity does not play a role in settling Sticky Situation. Instead, the role of this concept is to sift all norms into two categories: those that are merely formal and are not to be given serious consideration, and those that are authoritative. The issue of making a non-arbitrary selection between conflicting norms is only relevant for norms that are classified as authoritative. While the ought *simpliciter* picks out the correct final answer to the deliberation, the formal/authoritative distinction picks out which kinds of norms are competing in the first place.

The concept of authoritative normativity thus has a distinct role to play alongside the concept of ought *simpliciter*. At the same time, these concepts can inform each other. I will illustrate this in response to the gaps in our understanding of the ought *simpliciter* that Derek Baker highlights.[51] Baker argues that all accounts of the ought *simpliciter* either characterise it in other vague normative terms, which makes them circular, or reduce it to a non-normative property that does not seem to be what we are after.[52] I will focus on one such account, which he calls 'all-things-considered, reconsidered'. It is common to think of the ought *simpliciter* as an all-things-considered ought. It should be noted, however, that we can also talk of all-things-considered oughts within a system of norms: when I have moral reasons both for and against a course of action, the all-things-considered $\text{ought}_{\text{moral}}$ settles what, all moral things considered, I ought to do. In the context of the ought *simpliciter*, we are instead thinking of an all-things-considered ought that is not subscripted to a particular set of norms.

[50] I do not claim that McPherson conflates the two; however, the fact that 'authoritative normativity' is not uncommonly used to discuss the ought *simpliciter* signals that these topics tend to get mixed in the debate. In recent work, McPherson's paper is typically cited as if it concerns the formal/authoritative distinction (e.g. Plunkett, 'Robust Normativity, Morality, and Legal Positivism', 113).

[51] Baker, 'Skepticism About Ought Simpliciter'.

[52] I imagine that Baker would put McPherson's account in the first category: because McPherson leaves open what the success conditions of non-arbitrary selection are, his analysis of the ought *simpliciter* remains vague and in need of substance.

According to the all-things-considered account of the ought *simpliciter*, there is one single class of reasons, which moral reasons, prudential reasons, and so forth are subclasses of. Unlike subscripted oughts, the ought *simpliciter* relates to the complete set of reasons: "The all-things-considered ought (...) designates what all the reasons favor: it is the ought based on unrestricted consideration of the reasons."[53] On this view, all subscripted oughts are pro tanto, reflecting what would come out as the required course of action if you only had a certain subset of reasons. The all-things-considered ought, on the other hand, stands for what you actually ought to do considering all reasons, hence its authority.[54]

The problem Baker raises for this account is that there are prescriptive social conventions that do not seem to affect the all-things-considered ought, such as feudal norms and sexist etiquette rules. In Baker's terms, the reasons that determine the all-things-considered ought are ATC (all-things-considered) reasons. An advocate of the all-things-considered conception of the ought *simpliciter* then faces the following question: are reasons of etiquette and other conventions a subclass of ATC reasons, like moral and prudential reasons are?

The first option is to accept that all reasons are ATC reasons. As Baker shows, this results in counterintuitive first-order normative implications. If every system of norms produces a subset of ATC reasons, then arbitrary or harmful conventions have the same normative standing as morality and prudence. To illustrate, a mafioso's mafia reasons would partially determine his all-things-considered ought in just the same way as his moral reasons do. This result seems unacceptable. On that basis, Baker concludes that this cannot be the right conception of the ought *simpliciter*.[55] In addition, we can question whether there would even be any normativity left if we take all norms to provide reasons of the relevant kind for the all-things-considered ought. Spencer Case suggests that for every system of norms, an inverse system of norms exists that makes exactly the opposite demands. If all norms determine what we ought to do overall, then we have an unlimited number of reasons for and against everything.[56] Case

[53] Baker, 'Skepticism About Ought Simpliciter', 238.
[54] Baker, 238.
[55] Baker, 240.
[56] Linda Radzik makes a similar point in Linda Radzik, 'Justification and the Authority of Norms', *The Journal of Value Inquiry* 34 (2000): 458.

argues that this is tantamount to not having any reasons to do anything: in both cases, all courses of action are always normatively equal.[57] If this is right, accepting that all reasons are ATC reasons even results in normative nihilism.[58]

The second option is to restrict the types of reasons that figure in the all-things-considered ought. According to Baker, this is also problematic, as it does not allow us to treat etiquette as lacking authority: "The strategy we are considering characterizes being less authoritative in terms of being *prima facie*, of being based on a subclass of the ATC reasons. But the ought of etiquette is based on *all* of the (non-ATC) reasons of etiquette. It is not *prima facie* in any sense."[59] Our explanation of the authority of the ought *simpliciter* was that it is based on the full set of ATC reasons, rather than a subset of them. However, if etiquette reasons are not ATC reasons, they form their own class of reasons. Since the ought of etiquette is based on the full set of those kinds of reasons, not merely on a subset, we can no longer account for the ought *simpliciter* having more authority than the ought of etiquette. Baker emphasises that it does not help to claim that only an ought based on ATC reasons is authoritative; this is to characterise the authority of the all-things-considered ought in a circular way. Neither can we simply claim that ATC reasons are normative, whereas etiquette reasons are not; etiquette reasons show all hallmarks of normativity.[60]

Yet, there is a more promising way of characterising the normative difference between arbitrary conventional norms and norms that are involved in the all-things-considered ought. If we are able to give an account of the distinction between formal and authoritative normativity, this can be used to explain in what way etiquette reasons are less significant than ATC reasons: the former are merely formal, while the latter are authoritative. Only authoritative norms produce subsets of ATC reasons. In other words, the all-things-considered ought is determined by the requirements of

[57] Spencer Case, 'Normative Pluralism Worthy of the Name Is False', *Journal of Ethics and Social Philosophy* 11, no. 1 (2016): 6.

[58] This assumes that the reasons for and against something always have the same weight in this picture. This is plausible: if all reasons are ATC reasons, it is unclear what grounds there could be for ascribing more weight to moral than to immoral reasons. Reasons can have different weights, but a particular immoral reason would presumably have the same weight as the moral reason it is the inverse of.

[59] Baker, 'Skepticism About Ought Simpliciter', 239.

[60] Baker, 239–40. See Baker's quote above on page 32.

authoritative norms, while formal norms have no impact on it. Barry Maguire and Jack Woods treat this relation to the ought *simpliciter* as the defining feature of authoritative normativity,[61] as does Susanne Mantel: "qualified or standard-relative reasons are substantially [i.e. authoritatively] normative if and only if they bear on a comprehensive, unqualified ought together with reasons of all substantial domains."[62] On this picture, the ought *simpliciter* has a more significant kind of authority than the prudential ought because prudential reasons are only a subset of the ATC reasons. On the other hand, prudence still has a normative power that mafia rules lack, because the latter are only normative in the formal sense.

Regardless of the strength of this response to Baker, it highlights that there is a place for the concept of authoritative normativity next to the concept of the ought *simpliciter*. Because they both concern oughts with special authority, they are easily confused. However, it is useful to separate them, as they provide different pieces of the puzzle of normativity. The ought *simpliciter* has supreme authority, in virtue of being based on the whole set of authoritative reasons. However, there is also a less exclusive kind of authority that other oughts can have. The concept of authoritative normativity helps us to make a distinction between subscripted oughts that do and those that do not affect what we ought to do overall.

It is not necessary to combine a commitment to the formal/authoritative distinction with a commitment to the ought *simpliciter*: a sceptic about the ought *simpliciter* can still believe that some norms are authoritative, and vice versa. The formal/authoritative distinction can even be used to avoid an objection to rejecting the ought *simpliciter*. It may be thought that we need to posit an ought *simpliciter* because without it all coherent systems of norms would be on par. However, as long as we can still distinguish formal from authoritative norms, this worry is unfounded. At the same time, the two concepts can also enrich each other. The ought *simpliciter* accounts for what should be done when different authoritative norms conflict. And, as I illustrated, the concept of authoritative normativity may enable us to have a coherent understanding of the ought *simpliciter*.

---

[61] Maguire and Woods, 'The Game of Belief'.

[62] Mantel, 'Do Epistemic Reasons Bear on the Ought Simpliciter?', 218. In these accounts the concept of authoritative normativity is made sense of with the help of the concept of ought *simpliciter*, rather than the other way around.

This all assumes, of course, that a coherent and explanatory account of authoritative normativity can be found. Yet, there is some reason to doubt this. As I have shown, we cannot define authoritative normativity in terms of reasons, rationality, or other specific normative concepts, because these also allow for a formally normative interpretation. We are therefore left with rough characterisations and the vague idea that authoritative normativity has a special normative force that formal normativity lacks. Baker shows that similar descriptions are in use in the debate on the ought *simpliciter*, and points out that they are ultimately vacuous or questionable:

> talk of normative force is, to put it bluntly, completely metaphorical. The claim that the ought *simpliciter* is the ought that *really* tells you what you ought to do relies on the table-thumping sense of "really". (...) 'Overriding' and 'normative authority' rely on metaphors of political power, or they are simply vague ways of gesturing at some normative property which needs clearer characterization.[63]

Baker rightly makes us question if there truly is a particular property we are trying to grasp with our claims about force and authority. Yet, before we give up on a seemingly useful concept, we should continue to look for an acceptable account of the formal/authoritative distinction, while being cautious of the pitfalls Baker highlighted. The challenge is to define authoritative normativity in a way that does not change the subject, nor comes down to using other undefined or merely metaphorical normative terms.

4. Irreducibility

There is a lot at stake in how we conceive of authoritative normativity. When we use a demanding conception, it is more challenging to explain how any norm can be authoritative. The current debate gives fuel to scepticism about authority in this way. Discussions of authoritative normativity typically imply that the property of having

[63] Baker, 'Skepticism About Ought Simpliciter', 234.

authority must in some way be irreducible. At the same time, some claim that there are no irreducibly normative properties. In combination, these claims have the radical implication that there are no authoritative norms. Jonas Olson provides a prominent argument of this form for a normative error theory. In his view, a norm is authoritative if and only if we have a reason to comply with this norm which is not reducible to plain natural facts. Before explaining this, I will discuss an alternative interpretation of the feature of irreducibility, on which authoritative normativity is by definition not reducible to norms. Although this idea is less refined than Olson's, it will be useful to consider it first, as this will make clear what role norms do and do not play in the concept of authoritative normativity.

4.1 Irreducibility and norms

Olson defends an error theory about forms of normativity that are irreducible. While he does not discuss formal and authoritative normativity in these terms, it is clear from his description of irreducible normativity that he equates this with what I call authoritative normativity. He distinguishes two senses of 'normative': one that simply concerns correctness relative to norms, and one that involves irreducibly normative reasons. The former kind of normativity, which Olson calls 'reducible normativity', plainly fits the bill of what I call 'formal normativity'. Olson claims that the distinction he makes between reducible and irreducible normativity corresponds to Parfit's distinction between the rule-involving and reason-involving senses of normativity, as well as Anandi Hattiangadi's distinction between norm-relativity and normativity.[64] Regardless of the variation in terminology, it is clear from Parfit's and Hattiangadi's work that the distinction they are concerned with is that between formal and authoritative normativity. Overall, it is therefore safe to say that Olson takes his distinction between reducible and irreducible normativity to correspond to the distinction between formal and authoritative normativity.

Olson's error theory about authoritative normativity is based on scepticism about irreducibly normative favouring relations:

[64] Jonas Olson, *Moral Error Theory: History, Critique, Defence* (Oxford: Oxford University Press, 2014), 121.

> The irreducibly normative favouring relation is not reducible to an action's property of being a means to the satisfaction of some desire, or an action's property of being in accord with some rule or norm. When the irreducibly normative favouring relation obtains between some fact and some course of behaviour, that fact is an irreducibly normative reason to take this course of behaviour. Such irreducibly normative favouring relations appear metaphysically mysterious. How can there be such relations?[65]

One way to spell this out is to claim that a favouring relation is authoritative only if it cannot be reduced to an action's property of being in accordance with a desire or norm. This leaves us with the following picture of an instance of authoritative normativity: there is a fact that somehow favours a certain action, where the normative relation between this fact and this action is not explained by any desire or norm requiring this action. It is reasonable to be sceptical that such a favouring relation can exist.

However, this account of authoritative normativity results in the wrong categorisation of norms as authoritative or merely formal, even by Olson's lights. Within this first interpretation of the feature of irreducibility, there are two different ways to spell out the claim that authoritative normativity must be irreducible to norms. The first is that any favouring relation that is reducible to a norm is merely formal. The second is that a favouring relation is merely formal only if the reason involved is reducible to a norm. I argue that both fail: the first version entails that all norms – including moral norms – are merely formal, while the second entails that all norms are authoritative.

The first option captures the idea that authoritatively normative claims do not concern correctness conditions of norms. Nonetheless, it is a mistake to equate authority with not being reducible to norms. Consider moral normativity, which Olson sees as necessarily authoritative. When I have a moral reason to φ, the favouring relation between this reason and the action does not actually exist independently from norms; after all, it is a *moral* reason. The reason is related to the norms of morality.[66] The favouring relation between the fact that a child is drowning and the action of

[65] Olson, 136.

[66] My use of 'moral norms' should not be read to exclude moral theories that posit only a single moral norm, such as the norm to maximise overall wellbeing.

saving the child clearly arises in the moral domain. That an action is morally right just means that it is correct relative to the moral norms. Even in the case of moral normativity, then, the favouring relation does not mysteriously arise in a vacuum; it is directly connected to moral norms. Therefore, if being reducible to norms disqualifies a favouring relation from being authoritative, then morality needs to be moved to the merely formal category. This is incompatible with Olson's moral error theory: he argues that all moral claims are mistaken precisely because we are committed to moral reasons being authoritative.

The alternative option is to focus on the irreducibility of reasons. Here we need to distinguish between two forms reasons might take, which I will call norm-reducible reasons and substantive reasons. Norm-reducible reasons for φ-ing have the form 'φ-ing is correct relative to norm N', 'N requires φ-ing', or otherwise directly point to a norm. That something is morally wrong is a moral norm-reducible reason not to do it. Substantive reasons, on the other hand, are not directly reducible to facts about norms. In his discussion of irreducible reasons, Olson is focused on substantive reasons. While he claims that there necessarily are irreducibly normative reasons to φ if morality requires φ-ing, he denies that "if morality requires you to φ, then *that fact* is a reason for you to φ"[67] is necessarily true.[68] He takes moral reasons to have a substantive form: our supposed reason not to eat meat is not the fact that eating meat is wrong, but the fact that eating meat harms human and non-human wellbeing.[69] To get morality back to the authoritative category, we could therefore give the following specification of the idea that authoritative normativity must be irreducible: a favouring relation between a reason and an action can be authoritative despite being reducible to a norm, under the condition that this reason is not reducible to a fact about the norm. If moral reasons are substantive, they do not reduce to facts about moral norms, and would therefore come out as authoritative.

The problem is that we can say the same about other kinds of norms. We can understand the reasons of Catholicism, grammar, chess, and etiquette to have a substantive form that does not make reference to these norms. A Catholic's reason not to eat meat is the fact that it is Friday, rather than the fact that eating meat on a Friday is

[67] Here 'reason' must be read in the authoritative, not formal, sense.
[68] Olson, *Moral Error Theory*, 123.
[69] Olson, 118.

incorrect relative to Catholicism. The reason I have to capitalise 'Catholic' is that it is a proper noun. A chess player's reason to move their king is that the king is in check. When a child asks why they should not eat with their mouth open, 'You're in company!' is just as appropriate an answer as 'That's impolite!'. In all these cases, the immediate link to the relevant norm is clear: these substantive facts only function as reasons because the norms give them a normative meaning. The favouring relation between the fact that a man is indoors and the action of taking his hat off is connected to the etiquette norm that men must not wear a hat indoors. Yet, these reasons are not directly reducible to facts about these norms. It is possible to formulate reasons in a norm-reducible way, but this applies to moral reasons too. It seems arbitrary to conceive of moral reasons as substantive reasons while taking other kinds of reasons to have an explicitly reducible form. In the end, their structure is just the same: a fact favours a certain action in virtue of this action being correct relative to a norm when this fact obtains. Therefore, if reasons are authoritative as long as they have a non-reducible form, the category of authoritative norms becomes overcrowded: the reasons of grammar, chess, Catholicism, and even mafia norms are just as authoritative as moral reasons. This seems to be the wrong outcome. Moreover, it would make a normative error theory much wider than intended: scepticism about authoritative normativity now implies scepticism about all norms.

In conclusion, this potential understanding of irreducibility, on which it is claimed that authoritative favouring relations or reasons are irreducible to norms, fails. It leads to a conception of authority that is either so demanding that even morality does not have what it takes, or that is so insignificant that it applies to all norms. The difference between norm-reducible and substantive reasons is ultimately irrelevant; norms always play a central role, regardless of whether reasons wear these norms on their sleeves. We must therefore look for a different feature that separates formal from authoritative normativity.

The problems with the discussed account of authoritative normativity do point us in the direction of a more promising approach: since both formal and authoritative normativity involve a close connection to norms, our best chance of finding the real difference between the two seems to be to examine what it is that makes norms either authoritative or merely formal. When a norm is authoritative, it confers authority on the favouring relations, reasons, and requirements connected to it. When a norm is

merely formal, it still provides reasons, but these lack significance. An explanation of the authority of norms can therefore give us all we need. Our broad question about the nature of authoritative normativity now becomes clearer and more specific: what does it take for a norm to be authoritative rather than merely formal?

4.2 Irreducible authority

Olson's own view of irreducible normativity provides an answer to this question. He claims that a norm is authoritative if and only if our *reason to obey this norm* is irreducible. In debates on authoritative normativity, the crucial question is not what various norms require, but which norms we are required to obey: "To say that some behaviour is correct or incorrect according to some norm, N, is not to say anything normative. It is merely to say something about what kind of behaviour is required, recommended, or forbidden by N. *The normative question* (...) concerns whether there are irreducibly normative reasons to comply with N."[70] Here, a reason to comply with N is irreducibly normative when the favouring relation between the reason and complying with N is irreducibly normative. "It is not reducible, for example, to facts about what would promote satisfaction of [the agent]'s desires, or to facts about [the agent]'s roles or engagement in rule-governed activities."[71]

Olson is not alone in making it part of the definition of authoritative normativity that it cannot be derived from such natural facts; indeed, it is standardly assumed that for a norm to be authoritative, its authority must be intrinsic to it. Consider Preston Werner's definition of robust (i.e. authoritative) normativity: "An entity is *robustly normative* iff it is either fundamentally intrinsically binding (in the way that formally normative entities are not), or not fully explicable without reference to some fundamentally intrinsically binding entity."[72] Similarly, Derek Baker describes the distinction between formal and authoritative systems of norms as the distinction "between those that are inherently significant and those that are not."[73] Southwood implies that authority must be intrinsic when he contrasts the normative force of rationality with

[70] Olson, 120.
[71] Olson, 122.
[72] Werner, 'Why Conceptual Competence Won't Help the Non-Naturalist Epistemologist', 618.
[73] Baker, 'The Varieties of Normativity', 568.

that of grammar: "Local grammatical requirements are merely constitutive rules or conventions that do not possess any kind of intrinsic normative status."[74] In Enoch's view, for a discourse to be normative in the authoritative sense "it must be *necessarily* related to real reasons, it must have good normative credentials *in virtue of its very nature*".[75] Others specify that authoritative norms are genuinely binding on us irrespective of our desires or goals.[76] Finally, Parfit suggests that anyone who conceives of normativity as reducible must be talking about something other than the authoritative sense of normativity.[77] He claims that "[n]ormativity is either an illusion, or involves irreducibly normative facts."[78] Of course, opposing theories of normativity have been defended. However, in the more specific literature on the formal/authoritative distinction, these are hardly given the stage. With relatively few exceptions,[79] when authoritative normativity is discussed, there is an underlying assumption that the special normative power of authoritative norms must be irreducible and intrinsic to them.

Notice how much is excluded by this conception of authority. There is a wide variety of theories that take normative authority to be grounded in natural facts. Among these are theories beyond what is captured by 'naturalism', such as Kantian approaches. In attempting to explain the difference between formal and authoritative norms, we could appeal to their relation to our desires, deepest commitments, identity, or agency. A naturalist may want to say that having authority consists in promoting our preferences or fulfilling our human nature. Humean and Kantian constructivists hypothesise that authority arises from our contingent or necessary practical

[74] Southwood, 'Vindicating the Normativity of Rationality', 11.

[75] Enoch, 'Is General Jurisprudence Interesting?', 72.

[76] Terence Cuneo, 'Destabilizing the Error Theory', in *Epistemic Reasons, Norms and Goals*, ed. Martin Grajner and Pedro Schmechtig (Boston: De Gruyter, 2016), 74; Letsas, 'How to Argue for Law's Full-Blooded Normativity', 167; Mitchell N. Berman, 'Of Law and Other Artificial Normative Systems', in *Dimensions of Normativity: New Essays on Metaethics and Jurisprudence*, ed. David Plunkett, Scott Shapiro, and Kevin Toh (Oxford: Oxford University Press, 2019), 143.

[77] See especially pp. 439 – 444 and pp. 448 – 453 in Derek Parfit, *On What Matters: Volume Two* (Oxford: Oxford University Press, 2011).

[78] Parfit, 267.

[79] E.g. Maguire and Woods, 'The Game of Belief'; Errol Lord and Kurt Sylvan, 'Reasons: Wrong, Right, Normative, Fundamental', *Journal of Ethics and Social Philosophy* 15, no. 1 (2019): 43–74; Finlay, 'Defining Normativity'.

standpoint.[80] While such theories offer a broad range of interpretations of the nature of authoritative normativity, they have in common that they present authority as an unmysterious property directly related to facts about agents. Olson leaves no room for any of these accounts. In one fell swoop, he dismisses agents' desires, preferences, aims, roles, activities, and institutions as possible sources of reasons that separate authoritative from merely formal norms.[81]

When authoritative normativity is defined as being irreducible, nearly all metaphysically naturalist accounts of it are out of the question.[82] This may be precisely the purpose of such a definition: the claim that some domain is normative – to be understood in the authoritative sense – often functions as a premise in arguments against naturalist theories of this domain. This is not only a familiar strategy in metaethics, but is likewise used against naturalist reductions of belief, meaning, and so forth. Metaphysically naturalist theories of a normative domain are assumed to leave out the authoritative normativity involved in it. In the background of such arguments lies a particular conception of authority that excludes plain natural facts as a possible source of authority of norms. While the success of metaphysically naturalist theories can of course be questioned, the cards are stacked against them from the start: the standard conception of authoritative normativity is so restricted and demanding that it does not allow a large group of theories of authority to be seriously considered.

As a result, it falsely appears as if one cannot accept the existence of authoritative normativity without facing the challenges of non-naturalism. If authority is a property that is intrinsic to some norms and cannot be explained by any natural facts, this makes it fundamentally unlike the properties that populate our scientific worldview,

---

[80] Sharon Street, 'Constructivism About Reasons', in *Oxford Studies in Metaethics 3*, ed. Russ Shafer-Landau (Oxford: Oxford University Press, 2008), 207–45; Korsgaard, *The Sources of Normativity*.

[81] Elsewhere, Olson explicitly distinguishes between two kinds of norms with a different normative status: transcendent norms and immanent norms. In contrast to transcendent norms, immanent norms are "those whose reason-giving force depends on agents' engagement in certain goal-oriented or rule-governed activities or their occupation of certain roles, such as institutional or professional roles" (Jonas Olson, 'In Defense of Moral Error Theory', in *New Waves in Metaethics*, ed. Michael Brady (New York: Palgrave-Macmillan, 2011), 64).

[82] Olson allows that a naturalistic account of irreducible normativity is at least a hypothetical option (Jonas Olson, 'Error Theory in Metaethics', in *The Routledge Handbook of Metaethics*, ed. Tristram McPherson and David Plunkett (New York: Routledge, 2017), 64–65).

and at least in that sense mysterious. This view also brings epistemological problems: if the authority of norms is not grounded in facts about who we are, what we do or what we care about, it is unclear how we could have knowledge about which norms ought to be taken seriously. I will not develop a full argument against non-naturalism here.[83] Instead, I want to point out that because there are familiar metaphysical and epistemological worries about non-naturalism, there is some reason to be sceptical about the existence of authoritative normativity as long as we define this in an exclusively non-naturalist way. If you are committed to metaphysical naturalism, or for some other reason believe that there could be no non-natural normative properties, you are thereby pushed to follow Olson in concluding that the property of being authoritatively normative is not instantiated; there are only formally normative norms.[84] All of this depends on the standard substantive conception of authoritative normativity. It is therefore high time that we question it.

The standard conception implies that if there is an explanation of why a norm has normative significance, this significance could not be the same as authority. This is odd: what does authority have to do with a lack of explanation? Why can a norm only truly matter if nothing ultimately explains why it matters? To reiterate, the question under discussion is what it takes for norms to be authoritative. Answers to this could go in many different directions. One popular answer is that authority is an irreducible property that is intrinsic to some norms. An unfortunate result of this answer is that norms' authority goes unexplained: some norms simply do have it and others do not. We need to see this for what it is: an unsatisfying element of a particular theory. Instead, this lack of explanation has come to be treated as essential to that which we are theorising about: if you try to ground norms' authority in substantive facts about these norms, agents or the world, you are deemed to be conceptually mistaken.

To be more precise, the assumption that authority must be irreducible only leaves room for a limited kind of explanation of the authority of norms. While some of the proponents of the standard view claim that the normative status of norms must be inherent to them, non-naturalists can alternatively argue that norms' authority is

[83] I will elaborate on the challenges of non-naturalism in chapter four.

[84] Whether the sceptic concludes in line with Olson that moral requirements do not exist, or that they do exist but are merely formal, depends on assumptions about the concept of moral norms. I will cover this in chapter five.

derived from objective values which are themselves intrinsically and irreducibly normative. On this view, why certain norms matter is explained to an extent. Morality, for example, may be authoritative in virtue of its positive relation to wellbeing. The special normative status of wellbeing, however, is not explained. The commitment to irreducibility and intrinsic normative force ensures that the explanation runs out before authority is unravelled; in the end, it must be a brute fact that some norm or value is intrinsically authoritative, otherwise it would not yield real authority.

Understood in this way, the standard view implies that authoritative normativity cannot be explained in fully natural, non-normative terms. This is a contender for the right outcome of the debate on normativity – perhaps there are conclusive substantive arguments to the effect that you lose out on authority when you try to unravel it, or that natural facts are just too different from authoritatively normative facts.[85] Nonetheless, this position does need such substantive support; it is far too contentious and exhaustive to be the starting point of the debate. As a conceptual claim, it is unwarranted: irreducibility is not what makes authority authority. We can therefore at least attempt to give an explanation of authoritative normativity in natural terms without changing the subject. To insist otherwise, Olson, Parfit, and others must be using a narrower concept – that of irreducible authoritative normativity – which is not the one we are supposed to be concerned with.[86]

Instead of being required for it, irreducibility may even undermine authority. The question of which norms have authority is of direct practical importance. Christine Korsgaard suggests that non-naturalism is in the theoretical business of describing what the world is like and thereby fails to answer the normative question, which arises "in the heat of action".[87] When someone doubts whether a demanding requirement on them is truly obligatory, it is not sufficient to answer 'Yes, it does have authority,

[85] Parfit, *On What Matters: Volume Two*, 324–27.

[86] Stephen Finlay similarly suggests that non-naturalists and error theorists are using a different concept of normativity than their opponents – a concept which carries a particular understanding of authority and which does not reflect what ordinary normative judgements are about (Finlay, 'Defining Normativity', 210–11.) Sharon Street argues that Parfit uses an overly narrow concept of reasons in Sharon Street, 'Nothing "Really" Matters, but That's Not What Matters', in *Does Anything Really Matter?: Essays on Parfit on Objectivity*, ed. Peter Singer (Oxford: Oxford University Press, 2017), 121–48.

[87] Korsgaard, *The Sources of Normativity*, 91.

as a matter of fact'.[88] Furthermore, as Husi points out, it does not seem appropriate to choose which norms to obey "in virtue of some mysterious property norms possess, as if certain norms came attached with some quality-seal mandating acceptance."[89] We need to recognise a rationale for following norms that is substantive and not alien – to understand a norm's authority over our actions, we need to understand its relationship to us as agents.[90] A theory of authoritative normativity that starts from our preferences, practical standpoint or reflective nature is better suited to explain its practical importance. Even if a norm is authoritative in virtue of promoting some objective value, it is unclear why this should affect our actions, unless we have a connection to this value. Imagine that fashion norms happen to have a non-natural property of authority, completely independently of our attitudes on dressing well. Upon discovering this fact about the world, it still seems reasonable to ask – Why obey fashion? If authoritativeness cannot be analysed into a more familiar form of significance, it is puzzling how we could be bound by such an abstract property. A *sui generis* property that is not based in non-normative features of agents and the context of their actions seems like the wrong kind of thing to settle which norms we ought to follow. If that is what authority is, we can wonder if the fact that a norm is authoritative is as significant for us as it is supposed to be.

4.3 Categorising authority

It is not just non-naturalists and error theorists who, to their own convenience, employ a non-natural conception of authoritative normativity; even some naturalists assume that authority must be irreducible or intrinsic in some way. As a result, they take themselves to provide a theory of a normativity that falls short of being authoritative normativity. I will illustrate this by highlighting two such naturalist theories, starting with Evan Tiffany's deflationary normative pluralism.

Tiffany makes three main claims: (1) *contributory pluralism*: reasons have a variety of normative sources, including for example desires, social norms, and moral value; (2) *deliberative pluralism*: there is no neutral standard for weighing different types of

[88] Korsgaard, 37–40.
[89] Husi, 'Why Reasons Skepticism Is Not Self-Defeating', 432.
[90] Hampton, *The Authority of Reason*, 47.

reasons; and (3) *normative deflationism*: no reasons have ultimate authority.[91] Claims (2) and (3) are a denial of the existence of an ought *simpliciter* and of authoritative normativity respectively. When deliberating on what to do, we can take up different standpoints, which provide their own reasons-of-a-kind to act in a particular way. In Tiffany's view, this is all we have: there is no overarching standpoint or "Reason-as-such" which tells us what we ought to do all things considered.[92] No standpoint is the correct one. When the reasons provided by standpoints conflict, there is no objective fact of the matter which should be given more weight. However, there is a psychological fact about which reasons have authority for an agent: agents are disposed to treat some normative standpoints as more relevant than others in their deliberation. Thus, Tiffany does highlight a sense in which some norms, like the norms of morality, are more significant than others.

Why, then, does he present this as a denial of authoritative normativity, instead of as a deflated yet positive account of it? This is due to his specific conception of true authority: "the concept of genuine deliberative weight must refer to an objective property. ... This concept of an objective property underlying correct or 'veridical' deliberative authority I shall refer to as ultimate authority."[93] Tiffany does not assume that this property of authority is necessarily non-natural; indeed, he mentions various naturalist theories of it.[94] Still, the assumption that real authority is necessarily objective rules out considering a view like Tiffany's as offering an account of authoritative normativity, instead of something less.

Jack Woods provides a second example of a naturalist account of normativity that is presented as falling short of being authoritative. Instead of distinguishing only between formal and authoritative norms, Woods describes three categories of norms with increasing significance: standards of correctness, formal standards, and substantive standards. Woods uses 'standards' and 'substantive' where I use 'norms' and 'authoritative'; I will replace his terminology with mine for the sake of consistency. In his view, all norms are standards of correctness, but not all norms are formal norms. He

[91] Evan Tiffany, 'Deflationary Normative Pluralism', *Canadian Journal of Philosophy* 37, no. 5 (2007): 231–62.
[92] Tiffany, 240.
[93] Tiffany, 248.
[94] Tiffany, 249–50.

clarifies this distinction by pointing to the difference in importance between etiquette and schmetiquette, an arbitrary system of norms that has the same territory as etiquette but makes different demands.[95] Etiquette seems relevant for what we should do in a way that schmetiquette is not. Since etiquette is assumed to be merely formal, there must be a category of norms with a lower normative status than that. At the same time, Woods suggests that no norms have a higher status than formal normativity. However, this need not be a problem: he argues that formal normativity is more significant than has been assumed.

In Woods's view, formal norms have a significant normative power as a consequence of our community taking them seriously. When norms are accepted by our community, they are binding for us in an inescapable way. There is an important role for the appropriateness of sanctions here: "[S]tandards get their status as formal standards when violations are taken to license criticism and where we take this criticism seriously in the sense of being a sanction."[96] I cannot violate our etiquette norms without being liable to criticism, regardless of my reason for this violation. Woods argues that when a norm is part of our normative outlook in this way, we have a reason to do what it requires. Roughly, his suggestion is that accepting a norm involves desiring to comply with it, which then gives you reason to comply with it.[97] This, in turn, gives you reason to perform the particular actions required by this norm. Woods uses subscripted reasons strictly to refer to this kind of genuine but indirect reason: a $\text{reason}_{\text{etiquette}}$ to φ comes from our reason to comply with etiquette norms, which in turn stems from the fact that our community accepts etiquette as a guideline.

Woods thus argues that norms accepted by our community are binding, inescapable, reason-giving, that violating them makes one liable for sanctions, and that these norms are more normatively significant than other norms. Despite all of that, Woods still opts for calling these norms formal and not authoritative. This modesty is explained by a restrictive understanding of authoritative normativity. In Woods's view, norms are authoritative only if they are intrinsically reason-giving: "I view the best explication of substantiveness as holding that a standard δ is *substantive* insofar as ought_{δ}

[95] Woods, 'The Authority of Formality', 209.
[96] Woods, 220.
[97] Woods leaves room for alternative explanations compatible with non-desire-based views of reasons.

facts totally explain *generic reason* to do as δ demands."[98] By 'generic reason' he means unsubscripted reason. The generic reasons provided by an authoritative norm cannot be explained by something extrinsic, like an acceptance-based reason to obey this norm: "theorists should want morality to be substantive in a stronger sense; they should want the fact that morality says "ψ!" to explain, by itself, reason to ψ."[99]

In Woods's framework, the difference between formal and authoritative norms is subtle: they both provide reasons that deserve to be taken seriously, but the genealogy of these reasons diverges. The crucial difference between morality and etiquette is that while I have reason to φ just because I morally ought to φ, the fact that φ-ing is required by etiquette does not in itself give me reason to φ.[100] The explanation of reasons provided by formal norms has several steps: (1) my community accepts norm N, which gives me reason to comply with N; (2) having reason to comply with N gives me reason to perform the actions required by N; (3) N requires me to φ, therefore I have reason to φ. In contrast, step (3) constitutes the complete explanation of reasons provided by authoritative norms: if N is authoritative, I have reason to φ solely in virtue of the fact that N requires φ-ing. In the end, the distinctive feature of authoritative norms, as Woods conceives of them, is just that the explanation of reasons runs out quicker.

As I have highlighted, this understanding of authority is puzzling. It is odd that Woods seems to assume that what makes authoritative norms authoritative is that, in contrast to formal norms, their reason-givingness is unexplained. Yet, the outcome of his account hinges on it. He argues that the norms we take seriously are binding for us in a way we normally associate with authoritative normativity. He could have used this to argue for the triumphant conclusion that authoritative norms exist after all. Instead, he leaves us with the consolation that formal norms are all we need. Clearly, what we understand a naturalist metanormative theory to achieve depends on the details we – usually covertly – build into our concept of authoritative normativity.

It may be helpful to follow Woods in distinguishing three different kinds of norms. We can divide norms into the categories of those that lack significance, those that have a reducible kind of significance, and those that are significant in an irreducible way. The third category may be empty, if the non-natural property of intrinsic

[98] Woods, 'The Authority of Formality', 214.
[99] Woods, 215.
[100] Woods, 227.

normative significance does not exist. However, it is implausible that the second category is empty: as Woods shows, we can describe a metaphysically unproblematic form of significance that some but not all norms have. The standard view is that only the third category corresponds to the class of authoritative norms. My suggestion is that the bare concept of authoritative normativity is not restricted to this: we need to be open to the possibility that norms of the second category count as authoritative too. It may be thought that an irreducible or intrinsic form of authoritative normativity would be stronger or more special. However, this does not entail that norms whose normative power is reducible to natural facts cannot be authoritative. There is no need to define authoritative normativity as the most exclusive, objective, and powerful kind of normativity we can conceive of; this would make it a strange concept that is by definition extremely hard to fill. Instead, the essence of the concept lies in its contrast to formal normativity: authoritative norms have a special normative force that other norms lack. That this normative force is necessarily irreducible and intrinsic to norms is a substantive add-on to the basic concept, which needs to be defended.

5. Conclusion

In order to have fruitful debates on authoritative normativity, we need to start from an accurate, clear, and neutral understanding of this concept. The definitions of authoritative normativity that are most commonly used fail to stand up to scrutiny, which shows that we need to be more careful about how we characterise authority. It is also important to recognise that the formal/authoritative distinction is separate from the issue of the ought *simpliciter*. While the concepts of authoritative normativity and ought *simpliciter* are arguably closely related to each other, they do not stand or fall together. Furthermore, we must be critical of the standard view that authoritative normativity is not reducible to natural facts. Firstly, the idea that normativity is authoritative only if it is not reducible to norms fails. Secondly, the common assumption that the property of authority is necessarily irreducible and intrinsic to norms is not only unwarranted, but also implausible. Having a significance that is ultimately unexplained does not seem to be what makes norms authoritative. There is a substantive debate to be had about the nature of authority, in which the non-naturalist could

turn out to be right; however, the point is that the feature of irreducibility is not inherent to the bare concept of authoritative normativity. Therefore, metaphysically naturalist theories cannot legitimately be dismissed from the start.

Whether a metanormative theory is understood to account for formal or authoritative normativity is not a mere terminological issue. We tend to treat the normative status of norms as binary: they either do or do not matter. Mere formal normativity is associated with the norms that do not matter. Therefore, it is not sufficient to account for an especially significant form of formal normativity; anything short of authoritative normativity is supposed to be not good enough. We normally conceive of some norms, such as moral norms, as authoritative. If there is no authoritative normativity, we need to revise our understanding of these. Even if we can provide a theory of formal normativity that shows it to be more significant than we thought, scepticism about authoritative normativity remains a radical position that we should want to avoid. Therefore, it is crucial that we pay close attention to our concept of authoritative normativity and do not allow it to be burdened by unwarranted assumptions about the nature of authority.

CHAPTER 3

Sources of Authority

0. Introduction

In the previous chapter, I argued that the common conception of authoritative normativity is overly restrictive – when we focus on the essence of the formal/authoritative distinction, there is room for theories that do not present authority as an irreducible and intrinsic property. It is now time to elaborate on the forms such theories can take. Theories that avoid positing a *sui generis* property of authority altogether will be of particular interest, as these avoid the problems that support a normative error theory. Before I turn to the details of different metanormative theories, I will explicate a broad norm-centred approach to authoritative normativity. I can then categorise and discuss more specific theories through the lens of this approach. This chapter has three aims: (1) to explore how we can make sense of authoritative normativity with a focus on norms; (2) to display the wide variety of theories of authoritative normativity on offer beyond the view that authority is irreducible and intrinsic to norms; (3) to highlight the promise of metaphysically naturalist theories of authoritative normativity by showcasing different theories of that category. Rather than arguing for a specific theory, I am focused on broadening our perspective on which theories are available and owed serious consideration. I will point to the major advantages and disadvantages of

the different views, with the exception of Humean theories, objections to which I discuss in the next chapter.

I will start by presenting a framework for accounts of authoritative normativity that abstracts away from specific theories but shows the directions such theories can go in. In section 1, I will offer a rough diagnosis of the difference between authoritative and formal normativity which focuses on the justification of norms. This can then be specified and explained in a range of different ways. I will classify possible accounts of normative authority into three categories, according to how they conceive of the structure of justification of norms. This framework covers the common view I described in the previous chapter, but also contains explicit space for other theories.

In section 3, I will put a spotlight on several theories according to which norms are justified by a normative foundation with reducible authority. My aim is not to defend a particular view, but to show the diversity and appeal of theories that do not present authority as an irreducible and intrinsic property. In this context, I will discuss Christine Korsgaard's Kantian constructivism, Sharon Street's Humean constructivism, desire-based accounts as defended by Bernard Williams and Philippa Foot, and theories of authority offered by David Copp and Kate Manne which focus on the impact of norms on collective flourishing. Where these theories are not originally focused on norms and authority, I adjust their central claims to fit into my framework, which gives us a new version to consider. The theories I discuss have in common that they explain authoritative normativity in terms of natural facts about agents' commitments, making them compatible with metaphysical naturalism. They differ in which kind of commitments they identify as the source of authority. While each theory explains the relation between these commitments and normativity in its own way, one fundamental thought reappears throughout my discussion: the idea that normative authority is a force that arises out of commitments that are *ours*. This constitutes an important alternative to the understanding of authority that non-naturalism and normative error theory are based on.

1. Forms of justification

1.1 A norm-based account of authoritative normativity

The starting point for my framework for accounts of authoritative normativity is suggested by the discussed failure of the idea that normativity is authoritative only if it is not reducible to norms. We have seen that this view is false, because norms do not merely belong to the domain of formal normativity; even reasons standardly considered to be authoritative, such as moral and prudential reasons, are connected to norms. I (formally) ought to wear matching colours due to fashion norms, and I (authoritatively) ought to save the child from the pond due to moral norms. Formal and authoritative requirements, reasons, and oughts are related to norms in the same way. This suggests that the difference lies in a difference between these norms. We can say that the defining feature of authoritative normativity is that it is grounded in authoritative norms.[1]

Authoritative norms, in turn, are norms that have some special property that binds us to them. I will use the term 'justification' as a placeholder for this binding property. It is the case that we *really* have a reason to save the child from the pond because the moral norm requiring this is justified. On the other hand, if fashion norms lack justification, it is not a mistake to ignore what they require of you. They do not have the special property requiring us to obey them. In sum, the basic account of authoritative normativity I propose has two steps: (1) normativity is authoritative if and only if it is grounded in authoritative norms; (2) norms are authoritative if and only if they are justified.[2] This abstract account can be used as a basis for more specific theories. It leaves open what is required for norms to be justified. This is where specific theories come apart: there is room for a wide range of hypotheses on what type of justification authoritative norms have. With a focus on justification, we have a clear task before us: identifying the source of justification of authoritative norms. This is a more promising starting point for getting to the bottom of authoritative normativity. Because we can

[1] This is compatible with different views on how general (moral and other) norms are: reasons may come from abstract principles or highly particular prescriptions.

[2] This account is inspired by Linda Radzik's analysis of authority in terms of justified norms, e.g. in Radzik, 'Justification and the Authority of Norms'.

conceive of different forms and sources of justification, understanding authoritative norms as justified norms helps us to get a more expansive overview of the different kinds of theories of authoritative normativity to consider, including ones that tend to be overlooked.

Like with other normative terms, there is a sense of 'justification' that plays a role in formal normativity as well. When an action is correct relative to a certain norm, we can say that this norm justifies this action – that is, the action is justified relative to that norm. In the same sense, (systems of) norms can also justify complying with other (systems of) norms and themselves. Take David Copp's example of a pointless yet self-supporting norm: "a standard is to be subscribed to if it can be expressed in English in fewer than twenty words."[3] While this norm on what norms to accept applies to itself, it does not thereby make it the case that we ought to accept it. Not every kind of justification, then, is sufficient for a norm to be authoritative. We need to identify the right source of justification that gives the norms it justifies a more significant normative force than formal normativity. I will discuss three different accounts of the structure of the justification of norms: the first option is to view justification as intrinsic to the norms that have it, the second is to understand justification as a matter of coherence, and the third is to adopt a foundationalist account that identifies an ultimate underlying source of justification. The options I will discuss will not capture all metanormative theories; in particular, they will exclude non-cognitivist and hybrid views on the content of normative judgements. I will focus instead on a wide range of cognitivist theories. Nonetheless, it will be good to keep in mind that there are still more alternatives to prominent non-naturalist views of authority available.

1.2 Intrinsic justification

The first type of approach in my classification is that of the standard view of authoritative normativity I highlighted in the previous chapter. On this view, norms are authoritative if and only if they are intrinsically justified. Justification does not derive from an external source; it must be inherent to the norm in question. Having

[3] Copp, *Morality, Normativity, and Society*, 41.

authority is a matter of being intrinsically significant or binding.[4] Merely formal reasons, in contrast, are "relativized to rules or standards that need not be inherently normative".[5] If moral norms are authoritative, this is part of the nature of these norms themselves; it is not in virtue of their relation to agents or even to objective value. I have already discussed some of the problems with such an account. It presents it as a brute fact that certain norms are authoritative and others are not, leaving authority unexplained from the outset. As I argued in the previous chapter, if a norm's authority is not derived from some meaningful feature it has, it is less clear why this property determines what we ought to do. In addition, since norms fit into a naturalist worldview but norms with inherent authority arguably do not, the 'intrinsic justification' view bears the burden of the metaphysical and epistemological challenges of non-naturalism.

An additional problem for this approach is that it is inconsistent with a common way of thinking about normativity. While some systems of norms are treated as clearly involving or not involving authoritative normativity, the normative status of others is less clear. For these, a typical way to establish that they are indeed normative – in the authoritative sense – is to argue that they derive normativity from a system of norms considered to be authoritative, such as morality. This approach is standard, for one, among philosophers of law concerned with the authority of political obligations: "for almost everyone (...) who has pondered it, the problem of political obligation is a moral problem, and the obligation in question is a kind of moral obligation. To have a political obligation, then, is to have a moral duty to obey the law."[6] The thought is that laws have normative authority because morality has authority, and morality requires us to obey the law. Morality's authority, in turn, is sometimes argued to derive from the authority of a different system of norms. One may argue, for example, that we authoritatively ought to be moral because this is required by rationality.[7] Regardless of the truth of these proposals, it is important to see that we can make sense of the

[4] Baker, 'The Varieties of Normativity', 568; Werner, 'Why Conceptual Competence Won't Help the Non-Naturalist Epistemologist', 618.

[5] Benjamin Kiesewetter, 'Are Epistemic Reasons Normative?', *Noûs*, 2021, 13.

[6] Richard Dagger and David Lefkowitz, 'Political Obligation', in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta (Stanford University, spring 2021), https://plato.stanford.edu/archives/spr2021/entries/political-obligation/.

[7] E.g. David Gauthier, *Morals by Agreement* (Oxford: Clarendon Press, 1986).

idea of derivative authority they are built on.[8] It seems that if an authoritative system of norms requires us to obey another system of norms, requirements of the latter system are not merely formal, but authoritative. There is an important sense in which we ought to obey them. This is incompatible with the view that norms are authoritative only if they are intrinsically justified. In this way, two standard approaches to authoritative normativity cannot actually be combined, because the first says that authority must be intrinsic to norms and the second implies that authority can be derived from other norms. If we want to use the second approach, the first is in trouble.

There is a way for non-naturalists to avoid this problem: as I will discuss below, they can say that authority need not be inherent to a norm but can be derived from something else that is itself intrinsically authoritative. This gives room for derivate forms of authority while keeping the claim that it is ultimately irreducibly normative. Although authoritative normativity is often portrayed according to the intrinsic-justification view, as shown in the previous chapter, it is plausible that most non-naturalists actually prefer this alternative account.

1.3 Coherence

The second type of account of the source of justification, in contrast, focuses exclusively on justification derived from other norms. Linda Radzik has highlighted this option by drawing an analogy between the question of what justifies norms and the question of what makes beliefs justified.[9] A familiar type of theory on the epistemological side is coherentism: beliefs are justified in virtue of being part of a coherent web of beliefs. By analogy, we may understand authoritative norms to be norms justified in virtue of being part of a coherent set of accepted norms. This could be the set of norms accepted by a group, or by an individual.[10] To accept a norm is not to believe

[8] There is an alternative way to interpret these proposals: perhaps the law has no authority, but we nonetheless ought to act in accordance with it, because this behaviour is directly required by moral norms. I take it that this is not normally what we mean when we explain the normative force of the law in terms of morality; we want to say that the law, as its own system of norms, has authority.

[9] Radzik, 'A Coherentist Theory of Normative Authority'.

[10] In David Copp's presentation of coherentism, the relevant set of norms is the one that would be accepted by the group or person in question in certain ideal circumstances (Copp, *Morality, Normativity, and Society*, 41). However, adding this qualification means that justification is no longer treated as a mere matter of coherence. The requirement that the coherent set of norms would be

that it exists or that it is justified, but to subscribe to it as a standard for behaviour and intend to conform to it.[11] The broad idea is that when accepted norms are supported by each other, in the sense that complying with a norm in this set is required or at least permitted by the other norms, they thereby justify each other, without any being intrinsically justified. This is a subjectivist theory: whether a certain norm is authoritative for an agent depends on what norms she or her group accepts.[12] Coherence is not treated as a reliable indicator that a norm is, for independent reasons, justified; it is what makes the norm justified.

Radzik specifies that the relevant norms are those an agent accepts and reflectively endorses. According to her Reflective Endorsement Coherentism, "a norm is authoritative for an agent if and only if it, and everything that is counted as a reason for it, is supported by norms that one would endorse upon reflection, where the process of reflection would itself be reflectively endorsed and ultimately undefeated."[13] A norm is ultimately undefeated if it is not prohibited by any norms that are part of the reflectively accepted set.[14] On this theory, justification is entirely internal to an agent's subjective set of norms: even the norm "One ought to accept norms that one would endorse upon reflection" is justified just because (and just in case) the agent endorses it upon reflection.[15] Radzik argues that the subjective and internal character justification has in her view helps to account for its normative importance. If you disobey an authoritative norm, you are violating one of your own commitments. And if you do not accept that the norms you accept upon reflection are justified, you seem to disrespect your power of judgement and distrust yourself as an agent.[16] Therefore, even if – or because – the authority of norms is not intrinsic to them, there is still a lot at stake in obeying authoritative norms.

---

accepted in an idealised situation seems to need an external kind of justification itself, which does not fit within coherentism (Radzik, 'A Coherentist Theory of Normative Authority', 39).

[11] Copp, *Morality, Normativity, and Society*, 19–22.

[12] Radzik, 'A Coherentist Theory of Normative Authority', 38.

[13] Radzik, 35.

[14] Radzik, 33.

[15] Radzik, 34.

[16] Radzik, 35–36.

David Wiggins can also be understood to defend a coherentist theory of normative authority.[17] He argues that we cannot have a morally neutral conception of reasons.[18] The question of which kinds of reasons count as good reasons can therefore only be answered from within a moral outlook. We act and judge in the context of an everyday morality that supports certain values and reasons for action. This morality determines, for those that live under it, which reasons are to be taken seriously.[19] Wiggins understands morality itself not as an abstract code imposed top-down on agents, but as a function of the concerns on which ordinary people actually act. Which reasons we ought to take seriously, then, is a matter of which reasons we do take seriously. For example, because those with our moral outlook act on a concern with solidarity, solidarity counts as a good consideration to act on. To liberally translate this into my own terms, the theory of authority Wiggins suggests is that a reason (or, presumably, a norm) is authoritative if and only if it is treated as a good reason (or norm) to act on within one's moral outlook. Reasons are justified by having a place in the web of accepted considerations of the moral community. Like Radzik, Wiggins claims that we can only question and compare reasons against the background of what we already accept.[20] Yet, he emphasises that we are fully committed to respecting the concerns internally justified by our own moral outlook, describing it as "the sphere of the unforsakeable".[21] In the face of theories presenting reasons as objective or intrinsically authoritative, Wiggins aims to "dissipate the preconceptions that render needlessly mysterious the inward significance or internal aim of the norms that everyday morality so unmysteriously proposes to us."[22]

Whether a coherentist theory of authority can be successful depends on what the point of a justification of norms is supposed to be. For Radzik and Wiggins, authoritative norms are justified within an agent's or group's own normative context, but not

[17] His theory could alternatively be classified as a form of Humean constructivism, discussed below, because it focuses on what follows from our normative judgements. The boundaries between the categories of theories are not always clear; as I will highlight, coherentism has important elements in common with other metaphysically naturalist theories.

[18] David Wiggins, *Ethics: Twelve Lectures on the Philosophy of Morality* (London: Penguin Books, 2006), 208–9, 265–66.

[19] Wiggins, 260–66.

[20] Wiggins, 260–61; Radzik, 'A Coherentist Theory of Normative Authority', 35.

[21] Wiggins, *Ethics*, 329.

[22] Wiggins, 263.

outside of it. They can explain to an agent why, given her set of reflectively endorsed norms or moral outlook, certain norms are authoritative for her. Why ought I obey the demands of prudence, while I can legitimately ignore the norms of fashion? Because this is implied by norms I am already committed to. This justification may be sufficient for the agent in question. However, it provides nothing to say to outsiders who do not yet accept the relevant norms. Since justification is internal to the web of accepted norms, there is no external justification for accepting this web of norms in the first place. Radzik admits that if someone does not reflectively endorse the norm of reflective endorsement that forms the centre of her account of justification, it is not the case that she ought to accept this norm. Such an agent stands outside of the context where justification takes place, and consequently no norms are justified or unjustified for her.[23]

Wiggins similarly faces a problem with justifying norms to persons who reject morality, who have a different moral outlook, or who have not yet stepped into one. One purpose a justification of norms may have is to show that it is warranted to raise new generations to respect certain norms and not others. Wiggins' internal form of justification fails here: while we have good reasons for adopting and obeying our morality, infants do not yet share our moral outlook, and so our good reasons do not count as such for them.

1.4 Foundations of justification

The analogy with epistemic justification provides us with an obvious alternative to coherentist theories: a foundationalist theory of the justification of norms. Foundationalist theories can be filled in in different ways. What they have in common is that they identify some ultimate foundation of normative justification: norms are not authoritative in virtue of their intrinsic justification or their place in a web of accepted norms, but in virtue of a source of justification external to them. Having a positive relation to this normative foundation is what constitutes the difference between

[23] Radzik, 'A Coherentist Theory of Normative Authority', 35. Radzik suggests that someone who does not accept the norm of reflective endorsement may not be an agent at all. She thereby ultimately explains the significance of normative authority in a similar way as Korsgaard, discussed below.

authoritative and formal norms. This relation may be indirect: a foundationalist theory can allow that a norm derives authority from another norm that is in its turn supported by the ultimate foundation of justification. There could also be a longer chain of norms deriving authority from derivatively authoritative norms, as long as there is a point at which we reach normative bedrock. The big challenge, of course, is to identify a plausible foundation of authority, and to explain why this has a fundamental normative significance, such that we ought to obey any norm that it directly or indirectly supports. There is an important distinction here between non-naturalist theories that point to a foundation with irreducible authority, and theories that identify a foundation whose authority is reduced to natural properties.

1.4.1 A FOUNDATION WITH IRREDUCIBLE AUTHORITY

Even non-naturalists need not accept that it is a brute fact that moral norms are authoritative and fashion norms are not; they can instead identify a property that moral norms have and fashion norms lack which is responsible for their difference in normative status. A plausible candidate is the property of promoting something objectively valuable. For example, perhaps the norms of morality are authoritative in virtue of promoting overall wellbeing, where wellbeing itself has a normatively significant value. Other norms that promote overall wellbeing would then also be authoritative. On this view, the difference between formal and authoritative norms is that there is objective value associated with the latter, which entails that we ought to respect them and are criticisable for not doing so.[24] This value forms the foundation of authoritative normativity. To grant authority to norms with the right relation to them, the values must themselves have a form of normative authority. On the non-naturalist picture, this authority is irreducible; it cannot be reduced to natural properties. It is at this stage that the explanation of authority runs out. In the end, a *sui generis* authoritative entity remains: not an inherently binding norm, but inherently binding objective value. This may be what some non-naturalists who seem to accept the 'intrinsic justification' account of authoritative norms really have in mind.

[24] Charles Côté-Bouchard and Clayton Littlejohn, 'Knowledge, Reasons, and Errors About Error Theory', in *Metaepistemology*, ed. Christos Kyriacou and Robin McKenna (Cham: Springer International Publishing, 2018), 153.

This foundationalist version of non-naturalism can be seen as more sophisticated, as it can give a substantive explanation of why certain norms have authority. And because it does not require authoritative norms to be intrinsically justified, it leaves room for derivative authority – to illustrate, the law may be authoritatively normative because morality requires us to obey the law, where morality is authoritative because it promotes irreducibly authoritative value.[25] Nonetheless, this view faces the familiar metaphysical and epistemological mysteriousness of non-naturalism. In addition, it shares the unattractive result that authority is ultimately unexplained: there is no answer as to why certain values and not others have authority, and we may wonder, as I have highlighted, why this alien property should take a central place in determining what we ought to do.

1.4.2 A FOUNDATION WITH REDUCIBLE AUTHORITY

The final foundationalist strategy is to identify a reducibly authoritative source of justification. Most naturalist theories of authoritative normativity, as well as theories not typically labelled as naturalism but compatible with metaphysical naturalism, fall in this subcategory. There is a wide range of things such theories may point to as the foundation of authority, including both mind-independent and mind-dependent options. An example of the first kind would be the view that norms are justified if they promote wellbeing, where – in contrast to the non-naturalist picture – wellbeing has an objective yet naturalistically explained form of value. I will focus, instead, on foundations of justifications that relate to what agents are like or care about: think of preferences, aims, desires, roles, institutions, activities or subjective values. I will use 'commitments' as shorthand for this non-exhaustive list. Commitments can have different positive relations to norms: we can desire to obey certain norms, compliance with norms can be a (sufficient or necessary) means to the satisfaction of our preferences or aims, and norms can have a constitutive role in activities or institutions, in the way chess rules are constitutive of chess. According to this class of views, norms are

[25] Another possible view is that norms are authoritative if and only if they have intrinsic authority *or* have authority derived from other norms with intrinsic authority. The law may be authoritative because morality requires us to obey it and morality is itself intrinsically authoritative. This option avoids the problem of derivative authority I raised for the 'intrinsic justification' view, but shares the problems I highlight for the 'foundation with irreducible authority' view.

authoritative if and only if they, or norms that derivatively justify them, have such a positive relation to the right kind of commitments. There are different ways to fill in 'the right kind' here: theories may focus on agents' actual commitments or those that are universal, necessary, or would be had by idealised agents. It is plausible that when certain commitments are significant in some way, the norms they have a positive relation to are thereby significant too. It is a mistake to ignore the prescriptions of a norm that our own (actual or privileged) commitments require us to care about. In this way, our commitments seem to make a difference to the normative status of norms. When a norm is justified by our commitments, it matters what behaviour or attitudes are correct relative to this norm. This may be all there is to authoritative normativity.

To avoid the disadvantages of non-naturalism, these theories must not only identify a natural source of justification, but also explain in naturalist terms why this source has the power to bestow authority on norms. If authoritative norms have a special normative force in virtue of their relation to significant commitments, we need a story of what makes these commitments significant. The answer cannot be that the commitments in question are intrinsically authoritative, as that leaves us with *sui generis* normative properties that this type of theory aims to avoid. It also will not do to say that these commitments are significant because a norm requires us to have them; this norm would then be the true foundation of justification, and its authority would be unexplained. We need an alternative interpretation of authority, which lets us understand our commitments as authoritative for us even though they are not intrinsically binding nor supported by some deeper normative foundation. This may be the biggest challenge that foundationalist theories compatible with metaphysical naturalism face. Here it is especially important to have in mind the bare concept of authoritative normativity, unburdened by unnecessary preconceptions: what we are trying to account for is a form of normativity that is more significant than formal normativity. We need to find a normatively relevant way in which some and not all norms are justified. Understood as such, it is certainly not out of the question that agents' commitments can have a naturalist-friendly significance that allows them to provide this justification. All foundationalist theories will need to come with an appropriate explanation for why the foundation they identify can function as the basis of normative authority, and it should not be assumed prior to a substantive debate that only non-naturalist theories can succeed at this.

Let's take stock. The abstract view that authoritative normativity is normativity grounded in justified norms is compatible with a range of theories on the source of justification of norms. One possible view is that norms' justification must be intrinsic to them, but there are also coherentist and foundationalist theories available. Theories which point to different structures of justification can share important elements – for example, foundationalist theories can also make use of the notion of intrinsic justification. At the same time, foundationalist and coherentist theories can share the idea that authoritative norms can justify other norms. However, the details of accounts of normative justification can differ in many ways, which leaves us with a wide range of theories of authoritative normativity to choose from. The 'intrinsic justification' view, which much normative scepticism seems based on, is far from the only option. Granted, other kinds of theories can spark their own branch of scepticism: an alternative way of arriving at a normative error theory is to argue that authoritative normativity must be grounded in a foundational source of justification, combined with the claim that there is no such foundation – for example, because a foundation must be justified by a deeper normative foundation, leading to a vicious regress of justification.[26] At the same time, recognising the variety of accounts of authority allows us to explore more potential routes out of scepticism. There are various kinds of theories of authoritative normativity available that do not posit mysterious irreducibly authoritative entities. Although each has its own challenges to overcome, as I have shown, they deserve more serious consideration than they often receive. Here, I will explore different theories belonging to one subcategory: that of foundationalist theories that identify a reducibly authoritative source of the justification of norms.

[26] Copp, *Morality, Normativity, and Society*, 37–42; Linda Radzik, 'A Normative Regress Problem', *American Philosophical Quarterly* 36, no. 1 (1999): 35–47.

2. Explaining authority

2.1 Practical identity

Christine Korsgaard provides a clear example of an alternative interpretation of normative authority that allows her to build a positive theory of normativity without non-naturalist elements.[27] To Korsgaard, the problem of normativity is the problem that arises for beings, like us, with a capacity for practical reflection. Because humans are not limited to acting on our inclinations automatically, we face a choice about our actions: from the first-person perspective of agents, it is unavoidable to distance ourselves from our desires and choose which to act on. And because we reflect on which action to perform, we need reasons to choose specific actions and not others. Our peculiar reflective nature, combined with the necessity to act in some way, thus gives rise to our need for a source of normativity. In Korsgaard's view, reflection functions at the same time as the solution to this problem: "We need reasons because our impulses must be able to withstand reflective scrutiny. We have reasons if they do."[28] As long as we have principles for reflection, allowing us to reflectively endorse certain desires, norms, and actions, we have what we need. The challenge that remains is to show the source of principles of reflection that are in some way correct for us. How do I decide what to endorse?[29] When the problem of normativity is understood as such, it is unnecessary and inappropriate to make assumptions about the metaphysics involved in its answer. As Korsgaard emphasises, the problem is practical: central to it is "reflection about what to do, not reflection about what is to be found in the normative part of the world."[30] It is far from clear that this must be solved by the existence of some intrinsically normative entities; to deny this is to rely on a different understanding of normativity and its role for us, which needs a defence.[31]

---

[27] Kantian constructivism is normally not considered to be a form of naturalism, and I do not intend to imply that it is – I discuss it as a theory that is compatible with metaphysical naturalism.
[28] Korsgaard, *The Sources of Normativity*, 93.
[29] Korsgaard, 97.
[30] Korsgaard, 116.
[31] I base my discussion on an interpretation of Korsgaard as defending a reductive metanormative theory. For alternative options, see Matthew Silverstein, 'Inescapability and Normativity', *Journal of Ethics and Social Philosophy* 6, no. 3 (2012): 1–27.

Korsgaard's own theory is that principles for reflection are provided by our practical identity. She follows Kant in arguing that normativity must come from laws we give ourselves. To have a free will, we must not only reflect on our inclinations, but do so on the basis of principles that are not merely externally imposed on us. Our nature requires that we act according to a law that gives us reasons, but, as autonomous beings, this law must come from within ourselves.[32] "When you deliberate, it is as if there were something over and above all of your desires, something which is *you*, and which *chooses* which desires to act on. This means that the principle or law by which you determine your actions is one that you regard as being expressive of *yourself*."[33] This brings Korsgaard to focus on practical identities, which she understands as descriptions under which you act and value yourself.[34] There are many different practical identities: one can be a human, woman, daughter, friend, Christian and student at the same time. Identities can change or lose their practical importance to you. In Korsgaard's view, all of these give rise to reasons and obligations. You can determine whether your inclinations are consistent with your identity, which provides you with a reason to endorse them. When your identity forbids an action, you are obligated not to do it.[35] Korsgaard explains that the force of these obligations lies in the fact that if we violate them – or violate them once too often – we lose our integrity and identity.[36]

With this, Korsgaard shows us the possibility of conceiving of our identity as the source of normative authority. Identity seems to be of deep practical importance: something goes wrong if your actions contradict who you are. Failing to live up to your own conception of yourself and breaking your integrity is a fundamental loss. This may be where the special importance of authoritative norms come from: when a norm is supported by your identity, there is a significant sense in which you ought to obey it. We can formulate a general identity-based view of authoritative normativity as follows: authoritative normativity is normativity grounded in authoritative norms, where norms are authoritative if and only if they are justified by our practical identity. There is room for different theories on how our identity justifies norms, and especially

[32] Korsgaard, *The Sources of Normativity*, 97–98.
[33] Korsgaard, 100.
[34] Korsgaard, 101.
[35] Korsgaard, 101, 120.
[36] Korsgaard, 102–3.

on what the relevant kind of identity is. These may but need not focus on an identity that agents necessarily have. Nor do we need to presuppose that the relevant identity is one you *ought* to have; the fact that this identity is yours already makes it in some sense significant for you. Perhaps this sense of significance is just what we need for authoritative normativity.

Korsgaard does famously supplement her identity-based theory of normativity with an element of necessity, by arguing that there is a particular practical identity no agent can escape. This is the identity of a moral being, which Korsgaard equates with the identity of a human being.[37] With the assumption that only identities can give us reasons in the background, the fact that we are reflective beings who need reasons to act makes it a practical necessity to accept some identity as reason-giving. To fail to do this would mean that one has no reason to act and to live, which, in Korsgaard's eyes, makes it impossible to act.[38] Her account of normativity is therefore ultimately underwritten by the unacceptability of its alternative, normative nihilism.[39] Being a moral being is the fundamental reason-giving identity that brings us into the realm of reasons, where we need to be as reflective beings. This identity is at the same time the source of moral obligations. An important idea here is that valuing your humanity requires valuing the humanity of others. This, in turn, is supposed to come with a broad framework of moral obligations towards others. The authority of these obligations is then argued to be directly explained by the inescapability of treating your humanity as a reason-giving identity.

Since our attention normally goes to Korsgaard's theory of moral normativity in particular, an interesting contrast with her theory of normativity in general easily goes unnoticed. While following the demands of morality is constitutive of our moral identity, compliance with other norms is constitutive of other identities. And while Korsgaard argues that our moral identity is necessary, she takes other practical identities to be relativised to agents' circumstances and social worlds.[40] Yet, in her view, this does not make these other identities normatively insignificant:

---

[37] Korsgaard, 120–21.

[38] Korsgaard, 123.

[39] Linda Radzik, 'Incorrigible Norms: Foundationalist Theories of Normative Authority', *The Southern Journal of Philosophy* 38, no. 4 (2000): 641–42.

[40] Korsgaard, *The Sources of Normativity*, 129.

> I do not take the argument to show that all obligations are moral, or that moral obligations always trump others. In fact the argument requires – and our nature requires – that we do have some more local and contingent identities, which provide us with most of our reasons to live and to act.[41]

Korsgaard's broader theory of normativity is therefore that we have reason to do what is required by our identities, whether the identity in question is necessary or not. Our necessary moral identity, however, does ground the normativity of the contingent ones. In Korsgaard's view, our humanity requires us to have and obey some practical identity. When you violate the obligations generated by a role you are committed to, you therefore "fail yourself as a human being, as well as failing in that role."[42] If you did not endorse your humanity as a reason-giving identity, other identities would also not be normative.[43] Korsgaard thus presents our moral identity as the foundation of authoritative normativity: the identities that justify our actions ultimately receive their authority from the moral identity we cannot help but endorse as authoritative. While our necessary identity requires us to commit to some practical identities, it leaves open which, apart from excluding those inconsistent with valuing humanity.[44] In this way, even Korsgaard's constructivism leaves substantial room for contingency in authoritative normativity. Which norms are authoritative may differ strongly between a father, Christian, and football fan on the one hand and someone who is committed to being a friend, activist, and feminist on the other, despite their shared humanity. Moral norms, however, are authoritative for all.

The necessity of morality's authority is arguably the most dubious element of Korsgaard's account. A crucial step is that she equates the moral identity with the human identity. And to conceive of yourself as having the human identity, in her account, merely seems to mean recognising that you are a being with a capacity for reflection, who needs identity-given reasons to act. However, how does acknowledging our humanity in this sense directly put us into "moral territory", as Korsgaard

[41] Korsgaard, 125.
[42] Korsgaard, 121.
[43] Korsgaard, 125.
[44] Korsgaard, 130.

suggests?[45] How can any substantive moral norms be justified by recognising that we need reasons to act? There are two options here. On the one hand, we can accept that to recognise one's moral identity is nothing more than to recognise that one needs identity-given reasons to act. In this case, the moral identity is highly formal and is not constituted by a commitment to substantive moral norms – recognising that you must act on *some* reasons does not entail recognising the authority of moral reasons.[46] Hence, the necessary authority of morality is not vindicated.[47] On the other hand, we can understand the moral identity to involve, as its name suggests, a commitment to morality being authoritative. If we necessarily have a moral identity, this may then be argued to vindicate the necessary authority of morality. Understood in this way, however, it is implausible that having a moral identity is necessitated by recognising your humanity. There is a significant gap between being human and being a moral being here. Even if we grant that our being human makes it necessary that we are governed by some identity, why could we not opt for an amoral one? In response, Korsgaard may argue that having a substantively moral identity is somehow constitutive of our humanity, but that would prompt the question why we should be human, in this narrow sense – we may opt for being just like humans except for this commitment to morality.[48] Overall, although the general identity-based view of authoritative normativity may have some promise, I take it that Korsgaard has not shown that there is a necessary identity which provides all agents with substantive authoritative reasons.

2.2 The practical standpoint

Sharon Street's theory of normativity is importantly similar to Korsgaard's, both being forms of constructivism. Yet, they come apart in various significant ways. Most

[45] Korsgaard, 121.

[46] The 'empty formalism' objection to Kant echoes here, although I am focused on the difficulty of fixing the authority of morality based on a highly formal identity, rather than the difficulty of fixing the content of morality based on a highly formal categorical imperative.

[47] Korsgaard describes valuing your own humanity as taking your humanity to be a reason-giving identity. Because this is another merely formal identity which does not seem to involve particular substantive commitments, it cannot help to put more substance into my first interpretation of the moral identity.

[48] David Enoch, 'Agency, Shmagency: Why Normativity Won't Come from What Is Constitutive of Action', *The Philosophical Review* 115, no. 2 (2006): 169–98.

notably, they have radically different implications for the objectivity of morality's authority: Street's Humean constructivism rejects the Kantian idea that there are necessary substantive normative commitments for agents, which makes the authority of any kind of norm dependent on contingent commitments. Furthermore, while Korsgaard and Street both argue that agents impose normativity on themselves, their arguments for this have quite distinct flavours. Korsgaard's take is that even if there were *sui generis* normative properties in the world, external to us, they would be the wrong kind of thing to give us reasons; being an autonomous agent necessitates acting only on laws we impose on ourselves. Street's theory, on the other hand, is motivated more by the thought that there is no external source of reasons to be found – metanormative realism is mistaken about what there is in the world. This leads her to the alternative theory that normativity comes from us. In Street's eyes, there is no value without valuers, and no reasons without beings taking themselves to have reasons. She explains the source of normativity in terms of practical standpoints. You occupy a practical standpoint whenever you make a normative judgement, that is, when you judge "that some things provide practical reasons, or are valuable, good, bad, required, worthwhile, and so on."[49]

In Street's account, two things are required for some judgements about an agent having reasons to be true. First, the agent needs to make normative judgements. This is easily, if not necessarily, satisfied. Humans are the kind of beings that value things and take certain considerations to speak in favour of certain actions. The next, crucial step is that there are standards of correctness constitutively entailed by any practical standpoint. Street also describes these standards as "rules of practical reason" constitutive of the attitude of valuing or normative judgement.[50] If one's attitude does not conform to these rules, it is not wrong or irrational, but does not count as a normative judgement at all. As a result, our normative judgements have entailments that we cannot avoid. Street's constructivism says that 'A has a reason to φ' is true as long as this is correct from the practical standpoint of A in combination with the standards of correctness constitutive of having such a standpoint. There are no reasons to φ full stop; there are only truths about a certain agent having a reason to φ. Here truth is

[49] Street, 'Constructivism About Reasons', 209.

[50] Sharon Street, 'What Is Constructivism in Ethics and Metaethics?', *Philosophy Compass* 5, no. 5 (2010): 373–74.

relativised to the agent's normative commitments, not the commitments of the speaker.[51] Truths about reasons arise once an agent makes normative judgements, from which other true normative judgements follow:[52]

> according to metaethical constructivism, a creature has no reasons until the moment it starts taking itself to have reasons, for until then, there are no standards of correctness determining which attributions of reasons to it are true and which are false. With the first making of a normative judgment, however, standards of correctness are legislated into existence: certain other normative judgments are now constitutively entailed, whether the creature realizes it or not, and the creature is now properly said to be making a mistake if it rejects these other judgments.[53]

The constitutive standards of correctness provide an element of objectivity and room for errors to Street's theory of normativity – it goes beyond the view that an agent has whatever reasons they take themselves to have. You can be mistaken in your beliefs about what you have reason to do. The most prominent example Street gives of a standard constitutive of the practical standpoint is the means-end principle. She claims that when you judge yourself to have conclusive reason to φ, you must, out of conceptual necessity, also judge yourself to have reason to take the necessary means to φ. If ψ-ing is necessary for φ-ing, you therefore have reason to ψ according to your own standpoint, even if you are not aware that it is a necessary means and that you have reason to do it.[54]

Street emphasises that the correctness of normative judgements need not be merely a matter of instrumental standards: "The standards constitutively legislated by our normative judgements are highly complex and varied in structure."[55] However, her other examples of such standards do not seem to have the same constitutive status as

[51] Street, 'Constructivism About Reasons', 224–25.
[52] Although Street offers a reductionist view of normative facts, she believes that the normative concepts are primitive and not reducible. Street, 'Nothing "Really" Matters, but That's Not What Matters', 125–26.
[53] Street, 'Constructivism About Reasons', 237.
[54] Street, 228–29.
[55] Street, 230.

the means-end principle. Street claims it is constitutive of making normative judgements that "someone who judges that only facts of kind *X* are reasons to *Y*, and who recognizes that *Z* is not a fact of kind *X*, cannot also (simultaneously, in full awareness) judge that *Z* is a reason to *Y*."[56] Similarly, when you judge that only things an agent has control over can be reasons for praise or blame, then you make a mistake when you blame someone for something that was out of their hands.[57] Yet, these restrictions on our judgements do not seem to be constitutive principles of practical reason; they are the result of particular normative judgements, rather than necessarily involved in making normative judgements at all. To illustrate with the quoted example, it is not constitutive of normative judgements that one cannot judge that Z is a reason to Y; this depends on the judgement that only facts of kind X are reasons to Y (and the fact that Z is not of kind X). The real constitutive principle in play here, as in the example about reasons for blame, seems to be a general requirement of consistency. Another constitutive standard Street offers, that you cannot simultaneously judge that something is and is not a reason for a particular action, is in line with this.[58]

The central role of these standards sets Street's constructivism apart from coherentist accounts of normative authority. The right normative judgements are not whichever judgements cohere with what is already accepted; facts about reasons are constituted by conceptual entailments from the initial normative judgements. This relation is more strict and more linear than coherence: there is not a web of mutually supporting judgements, but a base of deepest judgements which necessitate layers of further judgements about reasons and value. Coherence is a necessary but not sufficient condition on these further judgements.

This base of normative judgements is itself not necessitated by anything but completely dependent on what the agent accepts. Since facts about reasons arise from whatever reasons the agent takes herself to have, everything is ultimately contingent in this account. The normative facts could have easily been different if you had started with a different practical standpoint, and will be different for other agents. An infamous implication of this is that agents with certain commitments have no reason to be moral: they make no mistake if they ignore morality's demands, because they

[56] Street, 229.
[57] Street, 230.
[58] Street, 229.

actually have no reason to obey it by their own lights. It is also not the case that, in a standpoint-independent sense, you ought not have such commitments. In Street's view, there is no such sense: "Nothing matters, ultimately, independently of the attitudes of beings who take things to matter. To matter is to matter from the point of view of someone."[59] We can only ask about the correctness of a normative judgement with (usually implicit) reference to a standpoint made up of other normative judgements.[60] When there are no further normative judgements to assess a judgement by, there are no standards of correctness in play. The initial judgements that will determine normative truths can therefore not be right or mistaken.[61] Korsgaard argues that if you are a being who makes normative judgements at all, particular judgements are unavoidable for you. Street denies this: it is only once you make particular judgements that other, related judgements become unavoidable for you due to the constitutive principles of practical reason. As a result, choosing one's basic normative commitments – and choosing to make normative judgements at all – is a radical choice, by necessity not made for a reason.[62] I will return to worries about the fundamental contingency involved in this kind of account in the next chapter.

As she describes it herself, Street's theory does not fit directly into my framework for accounts of authoritative normativity; she is concerned with explaining the truth of normative judgements and the existence of reasons, not with explaining the authority of norms. Still, her form of constructivism can easily be reshaped into a theory of the justification of norms that captures the spirit of her ideas. Street does not distinguish between formal and authoritative normativity; in order to fit her account into my framework, we should understand her idea that the practical standpoint is the origin of reasons as the idea that it is the origin of authoritative reasons specifically. We can likewise understand it to be the source of authority of norms. On Street's view, an agent's standpoint consists of a large web of normative judgements, with at its core the deep commitments the agent holds most closely. When your values and judgements about the reasons you have conflict, your core judgements settle what can stay

[59] Street, 'Nothing "Really" Matters, but That's Not What Matters', 121.
[60] Street, 'Constructivism About Reasons', 220.
[61] Street, 223.
[62] Street, 237–38.

and what must go.[63] It therefore seems to be specifically the core of your practical standpoint that ultimately determines which norms and reasons are authoritative for you. Next, we can reinterpret Street's constitutive standards of correctness as standards of justification: these principles tell us what relation a norm must (directly or indirectly) have to an agent's core commitments to be authoritative. For example, this authority-oriented version of the means-end principle entails that if complying with a norm is a necessary means to something the agent deeply values, this norm is authoritative. It is constitutive of valuing to take such a norm to be authoritative, and so it is authoritative, for this agent. In this way, we can understand "a set of interlocking, mutually supporting, and mutually consistent basic judgments about reasons held very deeply by you"[64] to be the foundation of authority.[65]

How can our practical standpoint account for the special significance of authoritative norms? To this question, Street can be understood to have a similar answer as Korsgaard. According to Street, our most strongly and centrally held values define who we are to an important extent. You can come to know who a person is by learning the values they have, especially the deepest ones.[66] If instead of valuing your friendships and taking yourself to have reason to be kind, you had valued solitude and cruelty, you would have been a different person. The normative significance of your commitments thus comes from the fact that these are *your* commitments. To violate the norms that are authoritative by the lights of your own practical standpoint is to act at odds with your own identity. In Street's eyes, the normative significance of this is not undermined by the contingency of our commitments: "Contingencies have shaped what reasons we have in the same way that contingencies have shaped who we are. But now that we're here, we can shed our normative reasons no more easily than we can shed ourselves."[67] Humean constructivism entails that I could have been a person with different authoritative requirements – but, actually, I am not. Because my core

[63] Street, 234–35.
[64] Street, 243.
[65] In this quote, Street highlights the coherence among the core normative judgements. However, this does not make her theory a form of coherentism; as I have explained, the core judgements justify further judgements by strict entailment relations and not by mere coherence.
[66] Street, 'Constructivism About Reasons', 235.
[67] Street, 245.

commitments are what they are, certain norms are binding for me in a way I cannot shrug off.

2.3 Desires

An alternative Humean view identifies agents' desires or preferences as the source of normative authority. This likewise makes the authority of norms a subjective and contingent matter. The main difference between such a naturalist theory and Street's constructivism is the type of attitude focused on: in Street's view, the attitude of valuing or taking something to be a reason is distinct from the attitude of desiring. In addition, a standard desire-based account of normativity does not include claims about constitutive standards of correctness, although the means-end principle does have an important role. The straightforward idea is that an agent has an authoritative reason to φ if and only if φ-ing is a means to satisfying a desire the agent has. To adjust this to a norm-centred framework: a norm N is authoritative for an agent if and only if complying with N is a means to satisfying a desire the agent has. This is a broad view that can be adjusted and filled in in different ways, particularly when it comes to which desires count.

A prominent example of a more specific theory comes from Bernard Williams's defence of an internal conception of reasons. He identifies an agent's 'subjective motivational set' as the source of authoritative reasons, which includes not only desires in a narrow sense, but also "dispositions of evaluation, patterns of emotional reaction, personal loyalties, and various projects, as they may be abstractly called, embodying commitments of the agents."[68] Williams believes that the relevant motivational set is not necessarily the one the agent actually has; reasons can be provided by motivations the agent would have after undergoing rational deliberation (broadly conceived), and motivations the agent would have in absence of relevant false beliefs.[69] The crucial condition, however, is that this deliberation must start from the agent's actual motivational set. Note that there is no presumption that a subjective motivational set is

[68] Bernard Williams, 'Internal and External Reasons', in *Moral Luck: Philosophical Papers, 1973-1980* (Cambridge: Cambridge University Press, 1981), 105.
[69] Williams, 103–5; David Sobel, *From Valuing to Value: A Defense of Subjectivism* (Oxford University Press, 2016), 121–28.

self-centred; an agent's other-regarding desires count just as much.[70] In general, the desire-based view of normativity is not a form of egoism.[71]

Why should we understand desires – in a narrow or broad sense – to be the source of authority? In the first place, such a view is often motivated by an assumed restriction on the source of (authoritative) reasons. Williams claims that for something to be a reason for action, it needs to be the kind of thing that we can act on – a reason that can feature in an explanation of a particular action.[72] In his view, it is not sufficient if a reason would explain an agent's hypothetical actions given some motivational set they could have had; something can be a reason for an agent only if it can explain their actions given their real motivational set or the motivational set they would have after deliberation.[73] Williams therefore conceives of a close connection between reasons and agents' psychology: "nothing can explain an agent's (intentional) actions except something that motivates him so to act."[74] This is then coupled with the Humean idea that only desire-like states can motivate, not beliefs or rational processes on their own. As a result, all genuine reasons for action must ultimately be grounded in the agent's desires, broadly conceived. Indeed, Williams presents the link to the agent's motivational set primarily as a necessary condition on the existence of reasons, not a sufficient one.[75]

A second type of support for this kind of theory is the same as before: reasons grounded in our desires would be *our* reasons. The fact that a norm is supported by your own motivating commitments seems highly relevant for how you should respond to it. That desires are personal and dependent on an individual's psychology does not disqualify them from being the source of normative authority; Philippa Foot has suggested that this instead makes desires a more meaningful basis for reasons. She

---

[70] Williams, 'Internal and External Reasons', 105.

[71] Kate Manne even argues that desire-based reasons can come from desires held by others, not just our own. See Kate Manne, 'Democratizing Humeanism', in *Weighing Reasons*, ed. Errol Lord and Barry Maguire (Oxford: Oxford University Press, 2016), 123–40.

[72] Williams, 'Internal and External Reasons', 102.

[73] Copp, *Morality, Normativity, and Society*, 185–88.

[74] Williams, 'Internal and External Reasons', 107.

[75] Sophie-Grace Chappell and Nicholas Smyth, 'Bernard Williams', in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta (Stanford University, fall 2018), https://plato.stanford.edu/archives/fall2018/entries/williams-bernard/.

claims it is better to recognise that our reasons for obeying morality "depend on contingent human attitudes" and cherish such attitudes, rather than demand moral behaviour "with an alien 'ought'."[76] In Foot's view, when using an unsubscripted 'ought', we cannot say that an amoral person ought to care about moral aims, such as diminishing suffering. However, "we must start from the fact that some people do care about such things, and even devote their lives to them". Given their common aims, for these people morally required actions "are necessary, but only subjectively and conditionally necessary".[77] Even for moral norms, whose authority is widely presented as objective and necessary, we must seriously consider the possibility that authority grounded in desires or other individual commitments is sufficient. As Foot emphasises, this constitutes a solid and reliable foundation of morality in practice:

> We are apt to panic at the thought that we ourselves, or other people, might stop caring about the things we do care about (...). But it is interesting that the people of Leningrad were not struck by the thought that only the *contingent* fact that other citizens shared their loyalty and devotion to the city stood between them and the Germans during the terrible years of the siege. Perhaps we should be less troubled than we are by fear of defection from the moral cause; perhaps we should even have less reason to fear it if people thought of themselves as volunteers banded together to fight for liberty and justice and against inhumanity and oppression.[78]

2.4 Collective flourishing and constitutive aims

Explaining norms' authority with reference to the commitments of individuals is not the only option; norms may instead be justified by facts about the group or society an agent is a member of. I already discussed one such theory in the previous chapter: Jack

[76] Philippa Foot, 'Moral Beliefs', in *Virtues and Vices and Other Essays in Moral Philosophy* (Oxford: Oxford University Press, 2002), 130–31.
[77] Philippa Foot, 'Morality as a System of Hypothetical Imperatives', in *Virtues and Vices and Other Essays in Moral Philosophy* (Oxford: Oxford University Press, 2002), 169. (This section does not appear in other printed editions.)
[78] Foot, 167.

Woods's view is that a norm is normatively binding for you if it is accepted by your community and part of your collective normative outlook.[79] Another example is David Copp's account of the justification of moral norms. He argues that societies need to have a moral code to flourish. A shared moral code helps members of a society to avoid conflicts, coordinate expectations and cooperate.[80] If there is a particular moral code that would contribute the most towards the flourishing of the society that accepts it, "then that code is justified in a fully relevant and adequate way."[81] In Copp's view, the source of justification here is not individuals with an interest in living in a peaceful society, but the society itself. The normative status of moral norms is therefore not dependent on the characteristics or commitments of individual persons.[82]

Copp's full argument for this view relies on his particular conception of rational choice and of subscription to moral norms. He argues that moral norms are differentiated from other kinds of norms not by their content, but by our attitude towards them. In his eyes, part of what it means to subscribe to a norm as a *moral* norm is to desire it to be the social moral code in one's society. A social moral code is the set of moral norms that is culturally transmitted, socially enforced and generally subscribed to as a standard for behaviour by persons within a society.[83] Because moral norms' role in society is presented as essential to them, the relevant context for justification shifts from individuals to societies. Copp then offers the following thesis on the justification of moral norms: "A code is justified as a moral code in relation to a society just in case the society would be rationally required to select the code to serve in it as the social moral code, in preference to any alternative."[84] Copp defends an account of rational choice focused on satisfying the agent's needs and values, which he also applies to cases of choices by a society. This results in the following principle: "A society is rationally required to choose societal option x rather than y if, and only if, option x would best serve on balance both its values or stably endorsed preferences and its basic needs."[85]

---

[79] Woods, 'The Authority of Formality'.
[80] Copp, *Morality, Normativity, and Society*, 106.
[81] Copp, 107.
[82] Copp, 116.
[83] Copp, 74–102.
[84] Copp, 104. In the course of his discussion, Copp adds several small amendments to this thesis, which need not concern us here.
[85] Copp, 192.

On the plausible assumption that subscribing to a moral code serves a society's needs better than choosing not to have moral norms, any society will have a set of moral norms that is justified relative to it.

Copp ultimately explains morality's authority with an appeal to rationality – the fact that we need a moral code to flourish is relevant because, on his view, a society is rationally required to choose what it needs to flourish. A crucial background assumption here is that rationality is itself authoritative. The authority of rationality, however, is not explained. Copp's justification of norms therefore does not get to the bottom of where authority comes from.

Kate Manne also locates the justification of moral norms in the social sphere, with a similar focus on their being conducive to our flourishing. However, she develops this idea in a way that does not rely on a particular conception of rationality. Manne argues that social practices are a source of normativity: they can generate reasons – which we should interpret in the authoritative sense – to conform to the norms involved in this practice. The social practices she focuses on in particular are social relationships, such as friendship and marriage.[86] These relationships are partly constituted by norms, such as the norm to help your friend or spouse if they are in trouble.[87] Manne argues that when it meets certain conditions, a social practice counts as valid. In that case, its norms have "genuine, normative force".[88] Because the practice of marriage is valid, and it is a norm of marriage not to let one's spouse drown, the fact that she is your wife is a normative reason with genuine force to save your wife instead of a stranger.[89] Manne suggests that the normative force of a broad range of moral reasons – those concerning what we owe to each other – can be explained in this way.[90]

The important question is, then, under what conditions a social practice is valid. At this point, Manne adds a consequentialist element to her account: "social practices

[86] Note that marriage is understood as a personal partnership here, not as a legal institution.
[87] Kate Manne, 'On Being Social in Metaethics', in *Oxford Studies in Metaethics 8*, ed. Russ Shafer-Landau (Oxford: Oxford University Press, 2013), 53–55.
[88] Manne, 55.
[89] Manne, 60–61.
[90] Manne, 53. For her account of more basic moral demands, see Kate Manne, 'Locating Morality: Moral Imperatives as Bodily Imperatives', in *Oxford Studies in Metaethics 12*, ed. Russ Shafer-Landau (Oxford: Oxford University Press, 2017), 1–26.

are valid in so far as they are conducive to human flourishing at large."[91] The positive consequences of social practices are not the source of reasons to comply with them, but only the background condition for the reason-givingness of these practices themselves. Manne argues that this requirement for validity is a result of the constitutive aim of social practices. In her view, the point of interacting socially is "cooperating and collaborating with one's fellow human beings in such a way that life is liable to go better for everyone amongst us."[92] Social practices can therefore be assessed with respect to how they contribute to collective flourishing, not because flourishing has an independent normative significance, but because it is the aim of these practices themselves. When a social practice does not meet the standards constitutive of interpersonal interactions, Manne suggests, it is misguided from the point of view of our shared humanity.[93]

While Copp and Manne only provide an explanation of the normative status of (a subset of) moral norms, their accounts can be developed more generally. Copp could argue that other kinds of norms with an intrinsic social dimension are also justified for a society if and only if collectively subscribing to these norms would best serve the society's basic needs, values and stable preferences. When it comes to more private types of norms that lack an important role in society, justification could instead depend on being conducive to the satisfaction of an individual's needs and values. In this way, a complete account of the justification of norms may assign different sources of justification to different kinds of norms, depending on these norms' characters.

Manne's focus on the constitutive aim of social practices has interesting potential beyond an account of moral reasons. Many systems of norms are social practices, as Manne understands them. Her view implies that we can assess the validity of etiquette, fashion norms, chess rules, and mafia norms with respect to how conducive they are to constructive cooperation and collaboration with fellow humans. Note that we seem to get the desired result for mafia norms because we do not assess them relative to their own internal aims, which they are presumably conducive to, but relative to the aims of social practices in general. Apart from the direct implications for non-moral norms in the social sphere, the more abstract idea central to Manne's account is

[91] Manne, 'On Being Social in Metaethics', 69.
[92] Manne, 70.
[93] Manne, 70.

worth highlighting: a practice is valid, and thereby reason-giving, if it is (sufficiently) conducive to the constitutive aim of this kind of practice. To fit this into my own framework, we can focus on norms instead of practices, and understand constitutive aims to provide the conditions for authority rather than validity. We then arrive at the following account of authoritative normativity inspired by Manne: authoritative normativity is normativity grounded in authoritative norms, where norms are authoritative if and only if they are (sufficiently) conducive to the constitutive aim of the class of norms in question. On this view, the source of justification of a norm is the aim that is tied to this type of norm itself. It is plausible that different norms can have different constitutive aims and therefore need to meet different conditions to be authoritative. This account can be developed in various ways: more specific theories can fill in how many classes of norms there are, which norms fall into which, and what the respective constitutive aims of these classes are. A shared feature of these theories will be that the relevant aim of norms is not externally imposed, but given by what these norms are like and what role they have in our collective lives.

A worry about this approach is that it can easily justify too many norms. We can question Manne's idea that all social practices have the constitutive aim of contributing to everyone's flourishing. Do mafia rules not have their own, decidedly different aim? Some norms have evolved to oppress groups in society. In general, it is plausible that many norms are constituted by an aim which we, or at least those of us harmed by it, find objectionable. This means we will get counterintuitive results when we ascribe authority to norms just in case they succeed in living up to their own aims.

3. Conclusion

In this chapter, I have presented a norm-centred understanding of authoritative normativity. I started by proposing an abstract account that different types of theories of authority can start from. This abstract account states that (1) normativity is authoritative if and only if it is grounded in authoritative norms, and (2) norms are authoritative if and only if they are justified. I then distinguished three broad categories of views on the form of justification that makes norms authoritative: (1) authoritative norms are intrinsically justified; (2) authoritative norms are justified by coherence

with other accepted norms; (3) authoritative norms are justified by a certain foundation of authority. Within the third category, I made a distinction between theories that view this foundation as having irreducible authority and theories that reduce its authority to something natural. I do not claim that this is the only way to categorise theories of authoritative normativity, or even the most natural one. As I have shown, we can find important similarities between theories that I have separated, as well as significant differences between theories I have put in a shared category. However, I take my classification to be conducive to recognising the wide variety of theories available. It brings out the option of understanding norms' justification as a matter of coherence, which is not usually considered. It also emphasises that the 'intrinsic justification' view is not the only way to present authority as a non-natural property. Using this classification, I have discussed the different approaches theories take towards explaining normative authority, and highlighted their main advantages and disadvantages. Of course, there is a lot more to say about the comparative strength of these theories, and I have not established which account of the source of authority is correct. My aim, however, has been to make clear that there is a lot to consider beyond the view that authority is a *sui generis* irreducible property. When a certain way of conceiving of authoritative normativity is common, it is easy to miss that alternative accounts can be defended without changing the subject. As I have emphasised, there are different types of theories that do provide a serious alternative to non-naturalism about authoritative normativity. There is more than one way to answer what it takes for norms to have authority: we can disagree, without being conceptually confused, about what kind of justification authoritative norms have and where this justification comes from. Accounting for the existence of authoritative normativity does not need to involve explaining how there can be irreducibly normative properties.

In the second part of this chapter, I discussed several theories that take authoritative norms to be justified by a certain normative foundation with reducible authority. Instead of irreducibly normative properties, they appeal to natural facts about our identity, practical standpoint, desires or collective flourishing to account for the justification of norms. This source of justification is itself not presented as intrinsically and irreducibly authoritative, but as normatively significant in a natural and, typically, more personal sense. Of course, I have far from shown the full range of possible theories in this subclass: for example, I have not covered theories that explain authority in

terms of the commitments of idealised agents. Nonetheless, I hope that my discussion succeeds in giving a sense of both the variety and richness of theories of authoritative normativity that can be found on the other side of the spectrum from the 'intrinsic justification' view.

Despite the diversity of metaphysically naturalist theories of authoritative normativity, including coherentist theories, there are also important similarities between them. These theories are built on common themes, which are filled in in different ways. The first theme is that authoritative normativity arises out of a standpoint: there is no realm of independently existing authoritatively normative facts which agents are able to reach out to; instead, the authoritatively normative facts are internal to the context of agents. Without anyone occupying the relevant position, there is no authoritative normativity. Coherentist theories also present normativity as coming off the ground on the condition that a standpoint is occupied, where this means that an agent has accepted a web of normative judgements.

The second common theme is that authoritatively normative facts are explained by the agent's (or judge's) commitments, broadly conceived. It is clear that the different theories put various kinds of commitments at their centre. Korsgaard focuses on the identities we are committed to. In Street's account, it is all about the normative judgements we make. Wiggins's theory is very similar to Humean constructivism in this respect: the basis of normative truths are the normative claims we accept. In other Humean theories, however, authoritatively normative facts are derived from the desires, preferences or aims we have or would have after some reflective process. Copp and Manne show us that we can alternatively focus on the commitments at the societal level: we can understand authoritatively normative facts to be grounded in the needs of our society, or in the constitutive aims of the types of norms that have currency in our society. Although there is significant variation in the details, the focus on commitments gives these theories an important joint approach to normativity: authoritatively normative facts are not imposed on us top-down, as an alien force; they come from us, as personal demands that have a clear relevance to the first-person question of what we should do. The normative status of authoritative norms is explained by their being tied to us.

The third shared element is the requirement of a form of consistency, which has the role of connecting our commitments to a host of authoritatively normative facts.

This element is not always emphasised, but is necessary to derive a substantial set of normative truths from our commitments. Radzik's focus on coherence in the justification of norms reflects this. Requiring coherence allows her to extrapolate from what we reflectively endorse – the authoritative norms are not just those norms that we actually accept, but also the norms coherent with these. Street achieves the same with the much stricter notion of constitutive entailment: although normative facts start with the making of normative judgements, they are not limited to the content of those judgements, since our judgements entail other true normative claims. The element of entailment that takes a central place in Street's theory ensures that we end up with a realm of normative truths that exceeds our actual judgements. Other Humean theories are less clear about the exact way in which normative facts are derived from our commitments, but usually seem to rely on the means-end principle: we ought to take the means to our ends. In combination with this requirement of practical consistency, our commitments can be understood to lead to authoritatively normative facts. I will come back to the role and status of this means-end principle in the next chapter.

An aspect that fundamentally divides metaphysically naturalist theories is whether the commitments that ground normative facts are themselves inescapable. Kantian and Humean theories come apart here. Humeans may ascribe a form of inescapability to their account: given the commitments we have, which may be very deep and impossible to shed without changing who we are, certain authoritatively normative facts are inescapable. Kantians, however, present at least some commitments that we start out with as inescapable – being a human being necessitates having them. This makes Kantian versions of commitment-based theories of normativity a lot more attractive to those with realist intuitions. On the Kantian version, authoritative normativity is still dependent on agents, arising out of our standpoint, but at least (some of) the authoritatively normative facts are the same for everyone, necessary, and in a sense objective. Out of all metaphysically naturalist theories, Kantian constructivism may be the one that non-natural realists would be most willing to accept – if it works. The main question is whether a Kantian theory does work. I have discussed the doubt that any substantial commitments are indeed inescapable. Moreover, it is noteworthy that Korsgaard claims that most of our reasons are provided by our contingent commitments. Even if we all share a core of essential commitments, it seems that we would need to appeal to our contingent commitments in addition to the inescapable ones to

account for the full realm of authoritative norms. It is therefore worthwhile to explore whether, in the face of common realist intuitions, we can account for authoritative normativity with a focus on contingent commitments.

The greatest challenge of theories of authoritative normativity is to explain why the source of justification identified has the power to bestow authority on norms. The justifier needs to have a special status itself. It may seem as if something can be the source of authority only if it is something we *ought* to be committed to. Non-naturalist theories capture this intuition head-on, by ascribing an irreducible authority to the source of justification. I have tried to show that metaphysically naturalist theories can also give a story of the special normative significance of the source of justification they identify: it is significant for us because it is, in some way, ours. The authority of norms comes from an identity we are bound to, a practical outlook that shapes us, or another kind of commitment without which we would not be the same. Rejecting such theories as mistaken – as normative error theorists must – requires providing substantial criticism of this alternative interpretation of authority without assuming from the outset that the non-naturalist take on authority is correct. In the next chapter, I will consider such objections to contingent-commitment-based accounts of authoritative normativity.

Chapter 4

Conventions, Contingency, and Crossing the Is/Ought Gap

0. Introduction

In the previous chapter, I showcased theories of authoritative normativity that build on natural facts about agents' commitments. I highlighted a distinction between Kantian theories, which take (at least some) authoritative norms to be justified by commitments agents necessarily have, and Humean theories, which explain the authority of norms in terms of contingent commitments. Now, I narrow my focus to Humean theories of authority. There is significant resistance to the idea that normative authority is based in commitments which we do not have necessarily nor ought to have. The common view that authority must be irreducible and intrinsic to norms does not come out of nowhere; it is partly motivated by the problems alternative accounts are thought to have. Different fundamental objections are raised to the general Humean approach. Even if the bare concept of authoritative normativity has room for Humean theories, it may be thought, these objections show that these theories fail. This result would make it more plausible that we must understand authority to be a non-natural property, which can in turn motivate a normative error theory. It is important, therefore, to examine how much traction these objections really have. The

aim of this chapter is to clarify the implications of the norm-based type of Humean theory I have presented, and to defend it against major objections.

By 'Humean theories' I mean theories that explain authoritative normativity as a matter of contingent commitments – that is, preferences, aims, desires, roles, institutions, activities or normative attitudes which agents have without necessity. Throughout the discussion, I will take two more particular theories as examples of this category. The first is Street's Humean constructivism, which focuses on contingent values or normative judgements. The second is a general desire-based view which takes inspiration from Williams. I am not concerned here with the details of Williams's actual view, such as how deliberation features in it; my intention is merely to explore the implications and strength of a theory that takes individual agents' contingent desires or preferences to be the basis of normative authority. I will leave open more generally whether it is agents' actual commitments that count or only those not based on false beliefs, for example. Furthermore, I will be focused on the norm-based version of Humean theories here, as presented in the previous chapter, rather than the theories as originally proposed. What I offer is a defence of Humean theories in the form they take in my own framework for theories of authoritative normativity.

I will start by illustrating how Humean theories explain the authority of norms – or the lack of it – in particular cases. An important upshot will be that conventional norms can be authoritative on a Humean view. This, I will argue, is a positive result. Next, I will tackle two broad fundamental objections to Humean theories of authority. The first is the problem of contingency. I will divide this into two objections and respond to each. The second broad objection is against the idea, implicit in Humean theories, that we can cross the divide between 'is' and 'ought' or between the natural and the normative. I will argue that we can question this divide. I will end with a discussion of the authority of the means-end principle. Along the way, I will flesh out my norm-based and commitment-focused picture of authoritative normativity.

1. Implications of a Humean account of authority

1.1 The theory in practice

Before we can assess the plausibility of theories that understand authoritative normativity to be derived from our contingent commitments, we need to be clear on how such theories account in practice for the authority of particular systems of norms. With that aim, I will illustrate what Street's constructivism and a general desire-based theory imply about the authority of morality, etiquette, and fashion norms. Since the verdicts of these theories depend on agents' commitments, I will sketch the outcomes for agents with different values and desires.

Let's start with the system of norms that contingent-commitment-based theories are thought to struggle to do justice to: morality. Unlike Kantian theories, they do not claim that everyone necessarily has commitments that ground the authority of morality. Yet, agents can, and normally do, have contingent commitments that make moral norms authoritative for them. The easiest case is an agent – call her Anna – who is committed to morality itself. She judges deeply and consistently that she ought to do what morality requires, and desires to be moral *de dicto*. This makes morality straightforwardly authoritative on both Street's and Williams's type of theory: for Street, it is authoritative because the judgement that it is is part of Anna's practical standpoint; on Williams's view, moral norms are authoritative for Anna because complying with them is a necessary means to satisfying the desire to comply with them. This result does not apply to an amoral agent, Brad, who truly does not have moral values or a desire to do what is right. The Humean can say that a person like Brad is abnormal, but not that he is impossible, and must accept that morality has no authority over him.

Now imagine a more multi-faceted agent, Chelsea. She does not consciously judge that she ought to do what morality requires, but she does value some of the goods that morality promotes – goods like sympathy, honesty, and a concern for others' wellbeing. Likewise, while she does not desire being moral for its own sake, she does have other desires – such as engaging with others with mutual respect – which moral norms are fit to help her achieve. Since complying with moral norms is a means to obtaining what she values and desires, morality is authoritative for Chelsea. This shows that Humeans do not need to rely on a direct commitment to morality to point to; values

or desires that are connected to moral norms are sufficient to explain morality's authority. To insist that morality lacks authority for Ben, we have to add that he does not have normal values and desires relating to others like Chelsea does. It is plausible that most of us are committed to morality indirectly, as a result of what it stands for and achieves, rather than as a moral-rule fetishist. Having no morally relevant commitments at all, in contrast, is highly unusual. Even for an extremely selfish person, doing what is morally required may be the best means to the satisfaction of their desires – as I will discuss in chapter six, cooperating with others under moral constraints arguably secures advantages for agents regardless of whether they have any concerns for others.[1] We would therefore need to fill in Brad's extended set of commitments in a very specific and peculiar way – presumably including strong anti-moral values – to preserve the outcome that morality has no authority for him.

Note that the indirect nature that our respect for morality usually has fits better with the Humean picture than with a non-naturalist take on authority. Our moral motivation plausibly comes from considerations about substantive and natural features of morally required actions, such as how they contribute to wellbeing. We do not usually obey moral norms because we recognise that they have an irreducible property of authority. If there is indeed such a property, it is not clear what role there is for it to play in our motivation.

Next, consider the normative status of etiquette. I will understand etiquette here as a broad system of culture-specific norms that concern acceptable behaviour in social settings – it includes rules about table setting and how to respond to a written invitation, but also greeting conventions and the expectation that one is quiet on a bus. Etiquette, too, could have authority resulting from direct commitments to it – Anna may simply judge that she ought to follow the etiquette norms, and desire to do so. Brad, on the other hand, ascribes no importance to etiquette norms and does not care about being polite. Assuming he also has no commitments that indirectly support etiquette, it is therefore not authoritative for him.

As with morality, the more realistic possibility is a third one. In Chelsea's eyes, etiquette norms are archaic and arbitrary constraints that one may as well ignore. However, she desires to live in a society that enables smooth and effective interactions, to

[1] Gauthier, *Morals by Agreement*.

be respected as a member of that society, and to show respect to others. She deeply values these things. As it turns out, etiquette has an important role in securing these goods: etiquette norms make specific solutions to coordination problems salient and create reliable expectations of others' behaviour in social contexts, which makes it easier to have successful interactions.[2] Complying with etiquette makes you appear polite and civil, and conveys to others that you respect them.[3] Based on this, the constructivist can say that Chelsea's actual judgement about etiquette is mistaken – her deeper normative judgements entail that she does have authoritative reasons to comply with etiquette norms. A desire-based view has the same result: since obeying the demands of etiquette is an (arguably necessary) means to achieving what she wants, etiquette is an authoritative system of norms for Chelsea. Again, agents like Brad are outliers – many of us do not consciously value compliance with etiquette itself, but if you have normal background preferences about the state of the social world and your place in it, then etiquette is authoritative for you.

Finally, we turn to fashion norms. I have been using fashion norms as an intuitive example of merely formal norms. The example is intuitive because fashion norms are clearly arbitrary, highly culturally-dependent, constantly changing, and do not appear to contribute to some higher good – all aspects that we tend to associate with norms that do not matter. Even this example, however, is too quick. Consider Anna again, who is not only dedicated to being moral and polite, but also cares a lot about being fashionable. She desires to wear only what is in style and takes herself to have reason to do so. The Humean account of authority implies that, in this case, fashion norms are not merely formal after all. If it is unfashionable to wear socks in sandals, Anna has an authoritative reason to wear her sandals sockless. Brad, on the other hand, has no interest in following the latest trends. Beyond achieving a minimal standard of acceptable clothing, he does not care about what he wears, nor does he judge that he ought to. Here we reach the expected result: fashion norms do not matter.

In this scenario, Brad's desires and judgements are not at all as unusual and alien as in the previous two. It is perfectly normal not to care about fashion, and – outside of social media – you would not be condemned for failing to adhere to its norms. We

---

[2] Cristina Bicchieri, *The Grammar of Society: The Nature and Dynamics of Social Norms* (New York: Cambridge University Press, 2006), 34–42.

[3] Cheshire Calhoun, 'The Virtue of Civility', *Philosophy & Public Affairs* 29, no. 3 (2000): 251–75.

can see the aim to be fashionable as an optional personal project that some persons take on, which may form an important part of their self-image. At the same time, we can also imagine an indirect justification for complying with fashion norms. Chelsea does not desire to have a fashionable wardrobe for its own sake, but she does wish to be noticed at parties and gain admirers on Instagram. She judges that she ought to look pretty and fit in with the cool crowd. For Chelsea, too, fashion norms are authoritative. Their authority need not depend on superficial desires – it can be argued that fashion does have an important function, for example, in building group cohesion or signalling people's roles. In any case, we can conceive of both direct and indirect desires and judgements which would entail, on a Humean view, that fashion norms are authoritative for a person.

My discussion of the authority of morality, etiquette and fashion assumes that these systems of norms have or lack justification as a whole. In contrast, one may think that an agent's commitments could justify some moral norms and not others. For example, we can imagine that Chelsea values honesty and some other goods related to morality but is not concerned with the suffering of remote strangers. Does this mean that moral requirements to prevent suffering of remote strangers are not authoritative for her? There are two ways to think of this issue. The first view is that morality gets justified as a package – as long as sufficient moral norms are sufficiently justified by an agent's commitments, moral norms are authoritative as a whole for this agent. While I am sympathetic to this approach, I will merely flag it here, as defending it requires an investigation of the metaphysics of (systems of) norms which falls beyond the scope of this thesis. The second possible view is that norms are justified on their own, which entails that not all moral norms are authoritative for Chelsea. In this case, the claim that morality has authority can be interpreted as the claim that a substantial majority of the norms that are labelled as moral have authority.

1.2 The authority of conventional norms

A significant upshot of these examples is that Humean accounts of authoritative normativity allow conventional norms to be authoritative. By conventional norms I mean "informal norms that emerge through the decentralized interaction of agents within a

collective and are not imposed or designed by an authority."[4] They are usually implicit, regulating our social interactions through the power of mutual expectations. Violations may be punished, but not through any formal procedures – informal sanctions can include being ostracised or gossiped about.[5] Conventional norms are contingent and often differ between cultures. Their subject matter can be small and mundane, as Karen Stohr illustrates:

> Consider how many unwritten rules govern an ordinary and comparatively simple activity, like buying a drink at a busy Starbucks. There are conventions about standing in line, deciding what to order in advance of one's turn, moving out of the way of other customers while waiting for one's caramel macchiato, making room for people at the sugar and cream station, not taking up an excessively large table, not taking up any table for too long a time, and cleaning up after oneself upon leaving.[6]

Our behaviour is constantly subject to a plethora of norms like this. And most of the time, we comply with them – regardless of the obvious conventional nature of these demands, we do take them seriously:

> The rules that govern behavior in my local Starbucks could be different than they are, and probably are different in other Starbucks locations around the world. But just about any place we go, some customs and conventions are operating and many of us feel ourselves bound to act in accordance with them, though without always being sure why.[7]

Much of the current metanormative literature suggests that our feeling of being bound by conventional norms is confused; they do not, in fact, have authority. When

---

[4] Bicchieri, *The Grammar of Society*, x. Bicchieri uses this as a description of social norms, which she distinguishes from conventions. She classifies etiquette norms under the latter category. I will not be concerned with this distinction here, and will use 'conventional norms' as a broader class including both social norms and conventions.
[5] Bicchieri, 8.
[6] Karen Stohr, *On Manners* (New York: Routledge, 2011), 2.
[7] Stohr, 3.

the contrast between formal and authoritative normativity is explained with the use of examples, the examples of formal norms are nearly always conventional norms, particularly norms of etiquette.[8] They are contrasted with norms that seem less culturally dependent, such as prudential and moral norms. This tradition, which seems to have been inspired by Foot,[9] suggests not only that all conventional norms are merely formal, but that this is obvious: it is treated not as an implication of a substantive account of authoritative normativity, but as a platitude to be captured by it. To illustrate, Mitchell Berman describes it as a 'hallmark' of artificial, human-made norms that they "do not confer 'real' normative force, and thus are not *really* binding – whatever, exactly, it *is* for a norm to be *really* binding".[10] Another example comes from Errol Lord and Kurt Sylvan: "Explaining the distinction between anemic and genuine normativity is a hard task, but there are clear examples on either side: norms of etiquette and grammar, on the one hand, and epistemic norms, prudential norms, and moral norms, on the other."[11] Despite how seriously we take them in our everyday actions, conventional norms are used as the archetype of mere formality. A substantial part of the literature on authoritative normativity gives the impression that the distinction between formal and authoritative norms just *is* the distinction between conventional and non-conventional norms.

This understanding of conventional norms makes sense on the standard view of authoritative normativity, according to which authoritative norms have an irreducible bindingness built into them. While this view does not directly entail that conventional norms cannot be authoritative, it does naturally suggest this. Norms with an intrinsic *sui generis* authority must be pretty special. This does not seem like the kind of norm we can simply create through social interactions and uphold with mutual expectations. We can fully explain the existence of conventional norms in natural terms, which does not seem compatible with them having a normative force that is not reducible to anything natural. Hence, if you accept this understanding of normative

[8] E.g. Mantel, 'Do Epistemic Reasons Bear on the Ought Simpliciter?', 217; Cuneo, 'Destabilizing the Error Theory', 77; Plunkett, 'Robust Normativity, Morality, and Legal Positivism', 113; Parfit, *On What Matters: Volume Two*, 267–68; Werner, 'Why Conceptual Competence Won't Help the Non-Naturalist Epistemologist', 617; Olson, *Moral Error Theory*, 120; Wodak, 'Mere Formalities'.
[9] Foot, 'Morality as a System of Hypothetical Imperatives', 1972.
[10] Berman, 'Of Law and Other Artificial Normative Systems', 141.
[11] Lord and Sylvan, 'Reasons: Wrong, Right, Normative, Fundamental', 51.

authority, you will likely also accept that there is a fundamental divide between authoritative norms and conventional norms, which cannot be bridged for specific persons with relevant judgements or preferences.[12]

This view does not reflect the space conventional norms take up in our normative landscape. While philosophers tend to focus on the demands of morality, the impact of conventional norms on our behaviour may be bigger: as illustrated by Stohr above, they regulate many details of our daily lives. And instead of treating them as normatively insignificant norms, we overwhelmingly comply with them. We expect others to follow conventional norms and judge or sanction them when they don't. Compare conventional norms that have currency in our society to an arbitrary set of rules I have just made up, which includes the norm that everyone ought to touch their nose three times before speaking. This norm applies to you and gives you a formal reason to touch your nose three times before speaking. Nonetheless, violating this norm is entirely insignificant. This is not the case for the conventional norms of your society: violating them is treated as a mistake, because they have a significance that my arbitrary rules lack.[13] Overall, at least many conventional norms seem binding to us. The claim that they are not actually authoritative – none of them, for anyone – is revisionary and contradicts our practical experience. It implies that we ought not *really* greet our neighbour with a wave or wear black to a funeral – but we feel strongly that we must do these things.

Even when philosophers go against the orthodoxy and argue that conventional norms are indeed normatively significant, they tend to build on the assumption that some norms are intrinsically authoritative and others – the conventional ones – are not. If conventional norms do have authority, this is argued to come from their relation to authoritative non-conventional norms, usually moral norms. Etiquette has been defended by arguing that it is part of morality: in Karen Stohr's view, "polite

[12] The common thought that conventional norms must be merely formally normative may also be partly explained by the idea that they are arbitrary in a way the norms of morality and prudence are not. The norms of etiquette could easily have been different. However, it is a mistake to confuse arbitrariness with a lack of authority – that etiquette's requirements could have been different from what they are does not entail that I am free to ignore them now.

[13] Woods makes this point with the contrast between schmetiquette and etiquette. Woods, 'The Authority of Formality', 209–10.

behavior is an extension of morality into small corners of our lives."[14] Cheshire Calhoun's alternative argument is that while being polite is not directly part of being moral, it does express your moral respect for others: "Civility requires obedience to social norms not for their own sake but for the sake of one important moral aim: the communication of moral attitudes to fellow inhabitants of our moral world."[15] George Letsas likewise argues that etiquette norms are normatively significant due to "moral considerations to do with respect and protecting reasonable expectations".[16]

In my view, this type of justification of conventional norms is dissatisfying. The first option depends on a very broad understanding of morality: mundane informal norms only fall under morality if you consider practically every norm about behaviour that affects others to be moral. The second option, on the other hand, cannot explain all cases of seemingly authoritative conventional norms. We can imagine cases where a norm of etiquette applies and appears binding, but the moral dimension has been removed because a violation of the norm does not communicate disrespect. Take the Dutch cultural norm of greeting acquaintances with three kisses on their cheeks. Many people dislike this norm yet feel obliged to comply.[17] Now imagine a person who hates these awkwardly intimate kisses so much that they refuse to greet people this way. Their visitors are aware of this and know that this person's refusal to engage in the usual greeting says nothing about their respect for them. Given this, there seems to be no moral harm here. Nonetheless, the greeting norm seems to have significant force for this person: although the satisfaction of their personal preference outweighs the negative aspect of violating the greeting norm, it is reasonable to imagine that they still feel they are doing something wrong. This means that the authority of conventional norms cannot always be explained by moral considerations.

Furthermore, the morality-based justification presents the authority of conventional norms as completely dependent on their link to moral normativity; they lack their own form of authority. However, many conventional norms seem to have a significance on their own, independently of their role in morality. Conventions have an

[14] Stohr, *On Manners*, 4.

[15] Calhoun, 'The Virtue of Civility', 273.

[16] Letsas, 'How to Argue for Law's Full-Blooded Normativity', 176.

[17] Interestingly, this norm was suspended during the pandemic for obvious reasons, but came back in force immediately when the policy of social distancing was dropped. I take this to be a sign of the strength of this type of norm.

important function in helping us to coordinate our behaviour and live together successfully. We seem to respect them for their own sake, and not merely for their relation to moral norms. A satisfying account of the authority of conventional norms will reflect their own significance, instead of deriving their significance from norms that are assumed to be intrinsically authoritative.

As I have shown, Humean views of authority have the capacity to do this.[18] When authority is not understood as an irreducible *sui generis* property, but as a natural property that is grounded in agents' commitments, conventional norms are not excluded from having it. The ontology of norms is now less relevant; the crucial factor is that they have the right relation to our commitments, which conventional norms can very well have. Norms of etiquette and fashion can be authoritative for an agent with a certain practical standpoint or set of desires. I take this to speak in favour of this type of view. In contrast to the standard view, Humean theories do not present the authority of types of norms as a black or white matter: it is not the case that conventional norms are either never authoritative or always authoritative for everyone. Instead, the authority of norms is directly related to agents, and varies with their commitments.

2. The problem of contingency

Applying a Humean theory of authority to particular systems of norms brings the fundamental contingency involved to light. Neither fashion, etiquette nor morality are necessarily authoritative; all can but need not have authority. On a Humean view, the normative force of a class of norms is not fixed. It is dependent on one's commitments, which are themselves taken to be not necessary but contingent. To get a verdict on the authority of a norm, we need to specify a particular agent this applies to – as we saw, the outcome can be very different for agents with different values or preferences. This element strikes many as a major flaw of Humean theories. These theories

[18] Manne, for example, suggests that systems of manners have normative force in virtue of their social benefit. Manne, 'On Being Social in Metaethics', 71.

contradict an objectivist view of norms' authority, here captured by Jean Hampton as applied to moral norms:

> the moral objectivist claims that the reasons given us by moral norms are reasons that we have necessarily, no matter the state of our psychology, our profession or interests, the views of our society, or our metaphysical or religious commitments. Whether or not we also have motives by virtue of knowing these reasons, whether or not they are decisive in the circumstances, and whether or not we have committed ourselves to them so thoroughly that we find it is impossible not to comply with them, these reasons direct us, or oblige us.[19]

I will discuss two forms a contingency-based objection to Humean theories can take.

2.1 Contingency and the concept of authority

One possible objection is that the concept of authoritative normativity leaves no room for contingency. A theory goes conceptually amiss when it presents authority as a contingent matter. A common response to presenting an account of authoritative normativity focused on contingent commitments can be captured as follows: 'If this makes authority contingent, then how is it still authority?' The assumption here is that it is a conceptual truth that the normative status of norms is not contingent but necessary. Given this, any theory that explains authority as dependent on contingent factors fails to be a theory *of authority* – it may explain a form of formal normativity, but it is ruled out that it successfully captures the authoritative kind of normativity we are focused on.

This objection echoes the dismissal of metaphysically naturalist accounts of authoritative normativity I dealt with in chapter two – there, the idea was that being intrinsic and irreducible is part of the concept of authoritative normativity, so that any theory that does not provide this falls short of being a theory of authoritative normativity. The response I gave there also applies here: we should not confuse a feature of popular theories of the subject matter with a necessary feature of the subject matter.

[19] Hampton, *The Authority of Reason*, 97.

Philosophers are used to thinking of authority as a property that norms necessarily have or do not have, as that is how it is presented in a broadly non-naturalist understanding of authoritative normativity. However, this does not mean that it is entirely out of the question that authority is contingent. As I have emphasised, we should focus on the core of the concept of authoritative normativity, which is that it is a more significant form of normativity than formal normativity. In my view, the claim that authoritative normativity is not contingent is another substantive add-on to the concept, which we can reject. As I discuss below, a Humean theory's unavoidable commitment to contingency is indeed relevant for debating the relative strength of this approach; however, it does not disqualify it as a theory of authority from the start.

One may object that this is too quick – does our normative language not prove that necessity really is part of the concept of normative authority that ordinary speakers use? When we present a norm as authoritative, we seem to do so with no regard to the listener's subjective commitments. You ought to fulfil your moral obligations, whoever you are. What prudence tells you to do matters, full stop. In the same way, we make blanket claims about norms not being authoritative – mafioso or not, mafia rules demanding violent behaviour should not be followed. Our judgements about which norms matter for what we should do do not appear to be relativised to agents with particular commitments, or to allow for exceptions for agents with a different standpoint. To take normative discourse at face value, it might be argued, we have to accept that applying necessarily – independently of contingent factors – really is part of the concept of normative authority. In that case, a Humean take on authority does not capture what we are actually talking about.

To avoid this result, we can try to interpret ordinary normative language in a way that does not ascribe a commitment to necessity – or anything else the Humean cannot offer, such as irreducible and intrinsic normative force – to the speaker. Stephen Finlay does just this. To understand normative claims, we must not only look at the literal meaning of what is asserted, but also take into account the context of the utterance, as well as the speaker's intentions.[20] Finlay argues that we can then legitimately interpret ordinary normative language to lack a commitment to absolute authority.

[20] Stephen Finlay, *Confusion of Tongues: A Theory of Normative Language* (Oxford: Oxford University Press, 2014), 116–44.

He admits that some parts of normative language, especially moral claims, have an element of inescapability. We make claims that imply agents cannot 'escape' moral prescriptions – it is appropriate to say 'You ought to φ' regardless of the hearer's ends. However, Finlay argues, this does not mean that moral claims are not in fact relativised to ends. When an ought-claim is used categorically, it is implicitly relativised to *the speaker's* ends.[21] We talk as if the hearer's attitudes are irrelevant because they are not the basis of our claim – our own attitudes are. The speaker's aim is not to state what the hearer ought to do given the latter's standpoint; it is to motivate the hearer to act in accordance with the speaker's end. The special force of prescriptions in a categorical form explains why we do not make the subjective preferences behind our ought-claims explicit.[22] Furthermore, leaving the end behind our claim implicit communicates that this end is objectively valuable or at least shared by our audience, which can be rhetorically useful. In Finlay's view, pragmatically expressing that your hearer shares your end when this is clearly not the case is a way of prescribing that the hearer adopts this end.[23] In this way, we can explain categorical normative language as a device for speakers to make their own ends salient and influence others to promote them.[24]

Although Finlay hereby defends a speaker-relativist view, which contrasts with the agent-relativism of traditional Humean views, we can build on his points to show how normative language is compatible with commitment-dependent views in general. Presenting a norm as necessarily having authority is more forceful than specifying that it has authority for agents with certain contingent commitments. Moreover, when you claim that others ought to follow a norm which is authoritative relative to particular commitments, you can thereby communicate that these commitments ought to be adopted. Assume, for example, that mafia rules are authoritative relative to a mafioso's values or preferences. It nonetheless seems appropriate to tell the mafioso that he ought not follow these rules. This need not mean that we believe that whether the mafia rules are authoritative is not contingent on one's commitments. Instead, inspired by Finlay, we can say that our claim 'You ought not follow the mafia rules'

[21] Finlay, 176–84.
[22] Finlay, 184.
[23] Finlay, 184–88.
[24] A challenge for this account is to explain why appealing to objective value and categorical obligations has a powerful effect when, as we claim here, ordinary users of normative language are not committed to a categorical and objective view of normativity.

expresses that the commitments we or most of us have do not support the mafia rules, and that the mafioso should have those same commitments.[25] Of course, much more is to be said on the best interpretation of this claim, as well as other claims that involve an apparent commitment to authority not being contingent. However, it is significant that Finlay shows an alternative interpretation to be available – we do not need to accept that our normative language is conclusive evidence against a Humean view, because we can plausibly understand the non-Humean elements as merely being used in normative claims as a powerful rhetorical device. The simple data on what kind of normative claims we make does not prove that necessity is indeed part of the concept of authority.

2.2 Implications of contingency

An alternative interpretation of the problem of contingency focuses on counterintuitive implications: Humean views of authority imply that norms which we consider to be universally and necessarily authoritative, without exceptions, are not in fact authoritative for everyone. Moral norms form the prime example. If Brad's values do not support being moral, this is supposed to show that there is something wrong with his values, not that morality has no force for him. The problem also runs in the other direction: whereas it seems that some norms are not and could not be authoritative, Humean views imply that they are authoritative for agents with relevant commitments. The strongest examples here are morally bad norms, such as mafia rules. I have argued that views involving contingency are not, as a conceptual matter, excluded from being a theory of authoritative normativity at all. Very well, the opponent may say, but this contingency does make them *bad* theories. We have a strong intuition that everyone is genuinely bound to their moral obligations, and that no one ought to commit violence required by mafia rules. Any theory that fails to capture this simply seems to get things wrong.

A first way of responding to this objection is to try to minimise the counterintuitive implications. To start, the Humean can admit that it is possible to imagine agents for whom seemingly authoritative norms have no force, but argue that such agents are

[25] Another possible explanation of the claim is that we are sceptical that the mafioso's deep values truly support the mafia rules.

very rare in reality, if they exist at all. It is important that, in the example I gave, morality came out as authoritative for the agent who does not care about being moral *de dicto* but does have a normal set of values and respect for others. We have to portray Brad as an extremely selfish person, to the point of being cartoonishly evil, to get the result that morality lacks authority for him. This illustrates that the counterintuitive implications occur at the fringes of the commitments agents might have, far away from what an average human cares about. A Humean view only seems to run into trouble once we arrive at extreme cases. According to Street, "the realist and the anti-realist can do an equally good job of capturing a huge swath of our intuitive judgments about cases, coming apart only with regard to esoteric ones imagined by philosophers".[26]

We need to understand what these esoteric cases are like to determine how problematic they are. Recall that Street's constructivism does not treat just any value an agent has as normatively significant; if a normative judgement you make is not in line with your core commitments, it is mistaken and does not affect which norms are authoritative for you. The counterexamples to her theory therefore have to be, as Street calls them, ideally coherent eccentrics – imagined agents who have extremely unusual values which are nonetheless fully consistent with their other values and the non-normative facts. These values can be self-destructive, oddly pointless or morally repugnant. The first type is illustrated by the ideally coherent anorexic,[27] the second by someone who only cares about counting blades of grass,[28] and the third by the Roman emperor Caligula, whose sole aim is to maximise the suffering of others.[29] We want to be able to make two claims about Caligula: (1) He makes a mistake in valuing the suffering of others. He has a value that he should not have. (2) Regardless of his immoral values, Caligula does not have an authoritative reason to maximise others' suffering;

[26] Street, 'Nothing "Really" Matters, but That's Not What Matters', 140.

[27] Allan Gibbard, *Wise Choices, Apt Feelings: A Theory of Normative Judgment* (Cambridge, Mass.: Harvard University Press, 1990), 171.

[28] John Rawls, *A Theory of Justice* (Cambridge, Mass.: Belknap Press, 1971), 432–33.

[29] Sharon Street, 'In Defense of Future Tuesday Indifference: Ideally Coherent Eccentrics and the Contingency of What Matters', *Philosophical Issues* 19 (2009): 273–98. The Caligula example is from Allan Gibbard, 'Morality as Consistency in Living: Korsgaard's Kantian Lectures', *Ethics* 110, no. 1 (1999): 145.

he ought (in the authoritative sense) to refrain from cruelty. Street accepts that her view does not support these claims.

However, she doubts that we can conceive of ideally coherent eccentrics existing in reality, once we imagine them in sufficient detail. Take the case of the ideally coherent anorexic, which is supposed to produce the counterintuitive result that, on Street's view, a person committed to anorexia has most reason to starve herself. For this case to work, Street shows, this person would have to value nothing more than starving herself to death in order to be thin, without any deeper commitments that require her to stay alive. This would make her a highly unusual person, unlike any real person with anorexia.[30] The imagined supremely cruel but coherent Caligula is unrealistic as well: according to Street, we should "err in the direction of assuming that an alleged amoralist is making a mistake, for in real life such a person is in all likelihood inconsistent, wrong about many of the non-normative facts, and self-deceived."[31] The real Caligula will presumably have had some form of concern for the wellbeing of some people, or at least selfish goals which can only be achieved by limiting his cruelty towards others. As evil as someone may seem, as long as we can plausibly ascribe basic human values and preferences to them, we can explain why they have authoritative reasons to be moral based on their own commitments.

What about mafiosi? Are they not real people for whom, on the Humean view, morally bad norms are authoritative? Since being in the mafia will be a major part of a mafioso's identity and way of life, we can assume that he values following the mafia rules. If he does not value this in itself, he will at least value goods – respect, financial rewards, safety – that he can only achieve by following these rules. However, the mafioso is also a human being, who will likely have important values that are at odds with the mafia rules – in particular, he likely cares to a significant extent about others' wellbeing and living in a peaceful society. Wodak describes a helpful real-life example of a pimp named Kevin. Kevin explains that because it was required by "the rules of the game", he violently punished a sex worker, Lois, for whom he felt much affection.[32] We can imagine that, despite feeling bound to the pimp norms, Kevin cares more deeply about Lois' wellbeing and being a decent person than about being a good pimp.

[30] Street, 'In Defense of Future Tuesday Indifference', 276–81.
[31] Street, 293–94.
[32] Wodak, 'Mere Formalities', 832.

If this is the case, and setting aside the difficulties of stepping out of harmful conventions one is wrapped up in, Kevin ought not follow the pimp norms. This shows that an agent's commitment to harmful norms does not entail, even on a Humean view, that this agent ought to follow these norms. The outcome depends on a full picture of the agent's commitments, which will usually include commitments that speak against following the harmful norms. The important question is which commitments are most deeply held by the agent, making them outweigh the others. It is possible that a mafioso's core commitments support morality and not the mafia norms. Another option is that they support both – moral norms are authoritative for this person but may not always win out over mafia requirements. For the especially ruthless mafioso with little interest in others, the outcome that mafia norms and not moral norms ought to be followed remains. Such a case may not be as unrealistic as Street suggests. Nonetheless, it requires specific and relatively unusual assumptions about the commitments of the agent in question.

Street does not merely highlight the rareness of ideally coherent eccentrics to minimise the counterintuitive results of her view in practice; her second point is that these eccentrics are so abnormal that our intuitions about them are not reliable. Our conviction that morality has authority for everyone is based on considering human beings like us, with a broadly normal range of commitments. The ideally coherent eccentrics are not like this – when we imagine them accurately, Street argues, we have to see them more as alien beings visiting from another planet than as human beings.[33] Our intuitions about what aliens have authoritative reasons to do are less clear and can easily be mistaken. Our theories of normativity are not built for them. It can therefore be argued that it is not a significant problem that a Humean view fails to capture our intuitions about which norms ideally coherent eccentrics ought to obey; instead, we can treat it as a neutral outcome that this view gives us a verdict on unusual cases where our intuitions may be misguided.

Even if the Humean cannot say that Brad has authoritative reasons to be moral, or that mafia rules lack authority for a deeply committed mafioso, there is a lot that can still be said about these persons. We can make stable and universal value judgements about norms even if these do not necessarily have or lack authority. Complying with

[33] Street, 'In Defense of Future Tuesday Indifference', 293–94.

morality is always what you morally ought to do, whereas following violent mafia norms is always morally wrong. Assuming that a Humean view does not require a relativist theory of the content of moral norms,[34] it is compatible with the standard idea that no one "escapes the requirements of ethics by having or not having particular interests or desires."[35] The negative evaluation of Brad as immoral therefore remains available. Other norms provide their own internal types of praise and blame: someone can be irrational, impolite or unfashionable regardless of whether the norms of rationality, etiquette and fashion are authoritative for them. The applicability of a norm's appraisal is independent of its authority, in just the same way that the question of whether an action is illegal does not depend on whether the laws in place are legitimate. As Woods emphasises, we can validly criticise someone – at least in a norm-relative sense – for violating a norm that they had good reason to violate: "rudeness is rudeness, even when warranted".[36]

It may be objected that norm-based appraisals are meaningless if the norm in question lacks force – we treat the accusation that someone is immoral as severe because we assume that morality is (universally, necessarily) authoritative. However, for most of us, it does have authority. Brad may be able to shrug off our verdict that his actions are morally wrong, but to us, this judgement is very meaningful. Our negative response to the immoral stands. In Street's words, "we can say that we loathe the ideally coherent Caligula; that it's awful, from our point of view, that he and his normative reasons are like this; and that the rest of us who do care about morality have every normative reason to lock him up, defend ourselves against him, and to try to change him if we can."[37] She suggests that we can even be justified in falsely asserting, to Caligula and to each other, that he does have most reason to refrain from cruelty.[38] As Finlay highlighted, such a false claim can function to make our own commitments salient and influence those not on board with morality.

---

[34] I return to this issue in the thesis conclusion.

[35] Foot, 'Morality as a System of Hypothetical Imperatives', 1972, 311.

[36] Woods, 'The Authority of Formality', 217. Woods makes this claim about formal standards that are in force for us, which he distinguishes from mere standards of correctness. See my discussion of his view in chapter two.

[37] Street, 'In Defense of Future Tuesday Indifference', 293.

[38] Street, 293.

After using different strategies to minimise the problem of contingency, a core of the problem remains. As a defence of the Humean approach to authority, I have suggested that persons like Brad are extremely rare, that our intuitions about his reasons are not reliable and that we can validly make all sorts of negative assessments about him. However, what we want to say in the end is that morality would have authority for this hypothetical Brad. This we have not achieved. The next strategy for the Humean is to admit that this is an unattractive result, but argue that it balances out against the problems their view avoids.[39] No one pretends that there is a flawless theory available. As undesirable as some implications of a contingent-commitment-based theory are, these need to be weighed against the problems of alternative theories. Contingency is the cost of vindicating authority while avoiding positing necessary commitments or non-natural properties. If we see the latter as more problematic than contingency, a Humean view can still come out as having the most "plausibility points".[40] Having raised doubts about the idea that there are inescapable substantive commitments in chapter three, I will now highlight three major challenges of non-natural theories of authority.

First off, non-naturalism comes with an epistemological puzzle. If normative properties are non-natural properties, which are often defined as being causally inert, how can we know about them? How could we have access to them? A theory of normativity that entails we do not have normative knowledge is unattractive. Therefore, the non-naturalist must either explain how our capacity for forming accurate representations of the natural world also enables us to reliably form true beliefs about the non-natural, or must posit a very special faculty for grasping the normative truth, such as a faculty of normative intuition.[41] In reality, our normative beliefs do not seem based on a special direct link to facts out there; instead, we can explain them as the result of purely natural factors, such as evolutionary pressures to accept particular norms and commitments. It would be a major coincidence if our normative beliefs, formed independently of the non-natural facts, successfully capture those facts. It therefore seems

[39] Street believes that we are forced to accept her theory despite its counterintuitive implications by "the untenability of realism plus the failure of Kantian versions of metaethical constructivism." Street, 'What Is Constructivism in Ethics and Metaethics?', 370.

[40] "Plausibility points" is taken from David Enoch, *Taking Morality Seriously: A Defense of Robust Realism* (Oxford: Oxford University Press, 2013).

[41] John L. Mackie, *Ethics: Inventing Right and Wrong* (London: Penguin Books, 1977), 38–39.

that if non-naturalism is true, our normative beliefs are not reliable.[42] A metaphysically naturalist theory does not face these problems: if authoritative normativity is grounded in agents' commitments, for example, it is no mystery how we could have access to facts about which norms are authoritatively normative.

A second challenge is to explain the supervenience of the normative on the natural. It is commonly accepted that when two things are identical in their natural properties, they must also be identical in their normative properties. There can be no normative difference in absence of a natural difference. On a metaphysically naturalist theory, it is clear why this is the case: the normative supervenes on the natural because normative properties are constituted by or grounded in natural properties. There could not be a difference in normative properties without a difference in the properties they depend on. If normative properties are non-natural, however, it is far from obvious why they must supervene on natural properties. The non-naturalist has two questions to answer: (1) Why do the normative properties supervene on natural properties at all? (2) Why do the normative properties supervene on the specific natural properties they in fact supervene on?[43] If no satisfying explanation can be given, this is a significant disadvantage for non-naturalism. Moreover, the fact that there is supervenience may directly undermine non-naturalism. Given that things with the same natural properties also have the same normative properties, we can identify a complex predicate, in strictly natural terms, which anything with a particular normative property necessarily satisfies. On the assumption that necessarily coextensive predicates must ascribe the same property, this means that normative properties must be identical to natural ones.[44]

Finally, there is an important metaphysical unattractiveness to non-naturalism. This can be captured by the broad objection that non-natural normative properties would be strange. As Mackie puts it, "If there were objective values, then they would be entities or qualities or relations of a very strange sort, utterly different from

[42] Sharon Street, 'A Darwinian Dilemma for Realist Theories of Value', *Philosophical Studies* 127, no. 1 (2006): 109–66.
[43] Enoch, *Taking Morality Seriously*, 136–50.
[44] Bart Streumer, *Unbelievable Errors: An Error Theory about All Normative Judgements* (Oxford: Oxford University Press, 2017), 9–41.

anything else in the universe."[45] One problem to take from this is that non-natural properties would be mysterious. We are used to the natural realm and have a reasonably good grasp of it. If normative properties are entirely separate from this, they are something puzzlingly unfamiliar and alien. Given the epistemological challenge, it is not clear how their nature could even become less mysterious to us. Theories that avoid this strangeness have a significant advantage. The second, closely related problem is that non-naturalism adds something to our picture of what exists. The contemporary worldview is a naturalist one – the background assumption to explaining the world is that it contains the kinds of things that the sciences study, and nothing beyond that. As a result, adding a non-natural realm is a major cost of a theory. It makes it incompatible with the accepted worldview. Of course, we need not simply take metaphysical naturalism for granted. Firstly, it is not actually clear how we should define the naturalistic worldview, and what, precisely, exists in it. Secondly, it is not beyond question that this is the worldview we should have, or that we should apply it to every domain including the normative. Hampton argues that, in spite of the practical and predictive success of our scientific theories, the "ontological faith" that "all that exists is what science (at any given time) posits" is undefended.[46] We should watch out for adopting scientism. Still, as a matter of neatness and simplicity, it makes a theory of normativity more attractive when it appeals to nothing over and above what we already accept in our general worldview. If specifically *sui generis* normative properties are not necessary for explaining anything, Ockham's razor requires us not to posit them.[47]

Furthermore, positing non-natural properties does not seem to help to capture the phenomenon of authority. As I have highlighted in the previous chapters, authoritative normativity is intimately connected to our own, practical question of what to do. As agents, we are bound to authoritative norms. It is unclear how non-natural properties, which are unlike anything else we are familiar with as humans, can explain this bindingness. Adding a non-natural realm to our naturalist picture does not solve the problem of normative authority, but adds a new mystery: how can something so alien be central to facts about what we ought to do? Radzik argues that, because the issue

[45] Mackie, *Ethics*, 38.
[46] Hampton, *The Authority of Reason*, 38–39.
[47] Cf. Enoch, *Taking Morality Seriously*, 50–84.

of which norms have authority directly concerns a first-person question about what to do, we need a justification of norms that makes sense from our perspective as agents.[48] "Whatever it is that makes morality binding on this agent had better be something that she can grasp and appreciate the import of, otherwise her question ['Do I really have to be moral?'] has not really been answered."[49] Hence, both the epistemological challenges and metaphysical oddness of non-natural normative properties raise serious doubts about how they can play the practical role authority is supposed to have. Considerations about agents' commitments, in contrast, both are accessible to the agent, and play a natural role in practical deliberation.

After this short discussion, my opponent may not be convinced that the costs of non-naturalism are worse than the cost of contingency that is attached to Humean views. Of course, there is a lot more to be said on whether non-naturalists can answer the challenges I have raised. Putting that aside, we run into a deeper issue here: philosophers have different intuitions on the relative importance of different merits or flaws of metanormative theories. While some are willing to avoid non-natural properties at the cost of contingency, others do not ascribe as much weight to the challenges of non-naturalism and are more concerned with capturing the apparent objectivity of normativity.[50] It seems to me that the debate reaches bedrock here. It is my own view that avoiding the epistemological and metaphysical problems of non-naturalism is worth facing the problem of contingency. However, I cannot convince my opponent that this is the case; they may have different preferences for which problems to avoid, and there is no neutral standard for metanormative theories to appeal to in order to show those preferences to be mistaken. Overall, all I can do and have done is to assess how big the problem of contingency truly is, point out how much of our intuitive understanding of authority Humean theories can capture – some of which non-naturalism misses out on – and remind those reluctant to accept a Humean theory that its unattractive features need to be balanced against the challenges of alternative views. There

[48] Radzik, 'A Normative Regress Problem', 40.
[49] Radzik, 'A Coherentist Theory of Normative Authority', 25–26.
[50] To illustrate, Enoch explains that he simply does not share his opponents' intuition that it is implausible that there are objective facts that demand certain behaviour of people regardless of their commitments. Enoch, *Taking Morality Seriously*, 136.

is a lot to be said for Humean theories even if their opponents' theoretical preferences make them unlikely to be convinced.

3. Pulling the rabbit out of the hat

A more fundamental objection to a Humean theory of authority is that it attempts to do the impossible. It may be thought that deriving authoritative normativity out of mere natural facts comes down to pulling a rabbit out of a hat. Unless the rabbit is not really authoritative normativity, or the hat contains hidden normative elements, this cannot be done. To respond to this worry, I will first discuss the is/ought gap and just-too-different objection, and will then explain how we can account for the authority of the means-end principle within Humean theories.

3.1 The is/ought gap

The is/ought problem seems to pit Hume himself against what I have called Humean theories. Claims about natural facts, such as facts about which commitments agents' have, are merely claims about what is. Claims about authoritative normativity, however, concern what ought to be. How can we jump from 'is' to 'ought' in this way? Hume famously complained about theories of morality that make this move: "For as this *ought*, or *ought not*, expresses some new relation or affirmation, 'tis necessary that it shou'd be observ'd and explain'd; and at the same time that a reason shou'd be given, for what seems altogether inconceivable, how this new relation can be a deduction from others, which are entirely different from it."[51] You cannot derive a normative conclusion from premises about matters of fact, unless there is an implicit normative premise. An important motivation for taking authoritative normativity to be irreducible is that it respects this gap: if you start from *sui generis* normative properties, you do not cross from 'is' to 'ought'. When you explain authoritative normativity as the product of norms and commitments, on the other hand, it does seem that you attempt to derive a normative conclusion from purely descriptive premises.

[51] David Hume, *A Treatise of Human Nature: A Critical Edition.*, ed. David Fate Norton and Mary J Norton (Oxford: Clarendon Press, 2014), 302.

Related to this, Parfit objects to naturalist metanormative theories that the normative is just too different from the natural. He claims that "normative and natural facts are in two quite different, non-overlapping categories."[52] Just like "[r]ivers could not be sonnets, experiences could not be stones, and justice could not be [...] the number 4", a fact about how we ought to act, in the authoritative sense, could not be identical to or consist in a natural fact.[53] In his view, a naturalist reduction of the normative is excluded on conceptual grounds – it is not possible that normative facts turn out to be natural facts, because the meaning of normative terms like 'reason' and 'ought' does not allow for this. This argument relies on a restrictive conception of authoritative normativity, which I have criticised in chapter two. At the same time, the popularity of this restrictive conception is arguably at least partly explained by the common thought that normative and natural facts are distinct. The straightforward idea that there is an unavoidable gap between facts and values, the natural and the normative, and 'is' and 'ought' puts a powerful limit on which theories are seen as viable. From this perspective, a reduction of norms' authority to contingent facts about agents appears to be a hopeless project.

In response, we can call the significance of the is/ought gap into question. Charles Pigden argues that the claim that you cannot derive an 'ought' from an 'is' is true, but only as a rather trivial logical thesis.[54] It is an instance of the more general principle that logic is conservative. As he understands it, the conservativeness of logic means that you cannot validly derive a conclusion which non-vacuously contains a certain predicate from premises that do not contain this predicate. A conclusion non-vacuously contains a predicate if this predicate cannot uniformly be replaced by any grammatically equivalent expression without affecting the truth-value of the conclusion.[55] An 'ought' cannot non-vacuously appear in the conclusion if there is no 'ought' in the premises. However, this is not a special feature of normative claims; Pigden highlights that, for just the same reason, you cannot validly derive a conclusion about hedgehogs from hedgehog-free premises. Yet, this does not establish that hedgehog facts are a

[52] Parfit, *On What Matters: Volume Two*, 324.
[53] Parfit, 324–25.
[54] Charles R. Pigden, 'Logic and the Autonomy of Ethics', *Australasian Journal of Philosophy* 67, no. 2 (1989): 127–51.
[55] Pigden, 133–34.

special kind of facts which cannot be reduced to natural non-hedgehog facts. The conservativeness of logic does not reveal any deep truths about how the normative relates to the non-normative. In particular, it does not show that normativity is a special property which is not reducible to something natural.

What remains is the intuition that the descriptive or natural and the normative are firmly separated in a substantive sense, beyond general logical limitations. This is undermined to some degree by recognising a formal form of normativity. It is widely accepted, including by opponents of naturalist approaches to authority, that formal normativity can be reduced to natural facts – it merely requires the existence of norms, which are, in themselves, not metaphysically suspect.[56] I argued in chapter two that formal normativity should indeed be considered a form of normativity, that it is prescriptive rather than merely descriptive, and that it involves requirements, reasons and oughts. This raises the question: which side of the is/ought gap does formal normativity fall on? For the reasons I just described, it does not fit squarely under the merely descriptive. Yet, if we place it on the normative side, the gap turns out to be easy to cross: facts about what we have formal reason to do directly reduce to descriptive, natural facts about the norms that apply to us. There is no fundamental division between non-normative claims and claims with subscripted oughts – 'I $\text{ought}_{\text{moral}}$ to keep my promise' says nothing more than that keeping my promise is required by the moral norms that apply to me.[57] When it comes to formal normativity, the just-too-different problem does not apply: facts about formal oughts are not entirely different from natural facts, but made up of them. On this basis, we can reject the claim that the normative and the natural are, in principle, distinct categories which never overlap. Similarly, the nature of formal normativity implies that there is no sharp distinction between the descriptive and the normative. This creates space to question the is/ought gap.

---

[56] E.g. Parfit, *On What Matters: Volume Two*, 308–9; Olson, *Moral Error Theory*, 121; Joyce, *The Myth of Morality*, 43.

[57] This is behind John Searle's demonstration of how to derive an 'ought' from an 'is' (John R. Searle, 'How to Derive "Ought" From "Is"', *The Philosophical Review* 73, no. 1 (1964): 43). He derives 'Jones ought to pay Smith five dollars' from 'Jones promised to pay Smith five dollars' with the help of an implicit premise that you ought to do what you promised. This implicit premise is warranted as a formally normative claim that describes what the institution of promising requires. As a result, the 'ought' in the conclusion is merely formally normative. Parfit also makes this point about Searle's argument (Parfit, *On What Matters: Volume Two*, 310–14).

Of course, the real challenge is to bridge the gap between 'is' and the *authoritative* 'ought'. Parfit specifies that it is normative facts in the authoritative sense that are too different from natural facts.[58] While it seems unproblematic to cross from descriptive facts to formally normative facts, the Humean needs to show that there is no insurmountable divide between descriptive facts and authoritatively normative facts. A good strategy here will be to show that we can get authoritative normativity out of formal normativity without adding anything that does not fit in the category of the natural. If formal normativity reduces to a natural property, and authoritative normativity is constituted by formal normativity and some non-mystical additional element, then authoritative normativity is not isolated in its own realm, separated from the natural. The challenge is then to show that although the element we add to formal normativity is natural, the resulting form of normativity is indeed authoritative. With this strategy in mind, I can flesh out the sketch of a norm-based account of authoritative normativity I offered in the previous chapter.

My picture of normativity takes inspiration from Finlay's analysis of the formal/authoritative distinction. He is not focused on the distinction between formally and authoritatively normative facts or properties, as I am, but instead aims to explain the difference between formally and authoritatively normative *judgements*. In his view, these judgements are about just the same kinds of facts and properties. The difference, however, is that authoritatively normative judgements are "*made from a relevantly motivated perspective*".[59] More specifically, someone's normative judgement is authoritative if and only if this judge desires the end or accepts the rule that the content of this judgement relates to. The content of these judgements, Finlay argues, is not psychological: normative claims are not claims about the subjective commitments of the speaker or relevant agent. Instead, psychological factors only have a role in classifying normative judgements as authoritative or merely formal.[60] On this picture, authoritatively normative judgements are understood as formally normative judgements backed up by commitments that relate positively to the content of the judgement.

Although the Humean accounts I am focused on diverge from Finlay's in significant respects, there is an important idea here to take on board: there is no fundamental

[58] Parfit, *On What Matters: Volume Two*, 324.
[59] Finlay, 'Defining Normativity', 207.
[60] Finlay, 207–10.

divide between formal and authoritative normativity.[61] Authoritative normativity is made up of formal normativity as the basis, plus a layer of significance provided by agents' psychologies. On a norm-based view, both forms of normativity come out of the requirements, reasons and oughts of norms. Authoritative normativity has more force because it adds something – according to a Humean version, it adds facts about agents' contingent commitments supporting those norms. I hereby use what Wodak has called the 'Hamburger Model' of authoritative normativity: authoritative normativity meets the minimal conditions for normativity and adds an extra feature, like a hamburger with cheese.[62] The advantage of this approach is that we do not attempt to establish normative authority out of nowhere. We start with a fully naturalist account of formal normativity grounded in facts about norms. The content of norms is established at this level – what the demands of etiquette are is not dependent on whether those demands are authoritative. Some norms – and their requirements, reasons, and oughts – stand out as significant due to their positive relation to agents' commitments. This provides, for these norms and these agents, a layer of authority over the formal normativity of these norms. On this picture, this is just what authoritative normativity is. What makes it true that I ought to wait in line? The fact that a norm which applies to me requires this. And why does this matter to me – in other words, why ought I *really* wait in line? Because my personal preferences or values commit me to the norm in question.

We can now question the is/ought gap even for authoritative normativity. I have given a story of normative authority with multiple steps, where no step clearly crosses an impossible line between 'is' and 'ought'. There is no point at which we pull a rabbit out of a hat. Deriving normativity from natural facts is not an insurmountable task – we do this by deriving formal normativity from norms. To account for authoritative

[61] One significant difference is that Finlay only explains a distinction between formally and authoritatively normative judgements here; he does not account for the existence of distinctly authoritative normative *properties.* A second difference is that Finlay's theory has a non-cognitivist element. He categorises his 'perspectivism' as a hybrid theory. It has a cognitive dimension – when it comes to what normative judgements are about – and a non-cognitive dimension – when it comes to the distinction between formally and authoritatively normative judgements. In contrast, the Humean theories I have focused on are purely cognitivist, giving subjective commitments a slightly different place in the story.

[62] Wodak, 'Mere Formalities', 832.

normativity, we then need to identify a source of significance that mere formal normativity lacks. Agents' commitments can plausibly provide such significance. The thought that there is a fundamental substantive – i.e. not merely logical – gap between the normative and non-normative assumes that normativity is a distinct, non-natural type of property. It does not prove that a naturalistic reduction is impossible, but instead presupposes it. In the same way, we can deny that the normative and the natural are in distinct categories that do not overlap. Enoch suggests that we reach another "dialectical impasse" here: many are on board with the just-too-different intuition, but most reductivists are not.[63] In his view, there is little to say to convince those with the opposite perspective.[64] However, there is a way to settle this issue after all: if we can develop an accurate reductive account of authority, this would show that the understanding of authority behind the just-too-different intuition is mistaken. Non-naturalists cannot simply keep asserting that there is a gap between the natural and the normative; they need to respond to theories, like the one I have sketched, that entail there is none. While the opponents of Humean theories may intuitively accept a non-naturalist understanding of normative authority, it is precisely that understanding we are questioning here, and pointing to non-naturalist intuitions will not suffice to establish that a Humean alternative cannot work.

3.2 The authority of the means-end principle

I will end with a tentative discussion of a remaining worry. I have said that agents' contingent commitments provide a kind of normative significance. For a Humean view to be plausible, this needs to be the *right* kind of significance. Being justified by one's commitments needs to be what elevates norms to authoritative norms. For this to be the case, there would have to be an important sense in which you ought to follow the norms that stand in the right relation to your commitments. Authoritative norms are binding – therefore, the natural facts that explain their authority must bring a type of practical necessity to complying with these norms' demands. In Humean theories, the means-end principle has a crucial role to play here. It forms the link between our commitments and authoritative norms: a norm is binding for us when compliance

[63] Enoch, *Taking Morality Seriously*, 107.
[64] Enoch, 108.

with it is required for satisfying our preferences or promoting our values. On the Humean picture, there are no commitments we ought to have or have by necessity. Yet, once we have a set of commitments, the means-end principle is supposed to make it, in some sense, practically necessary to obey the norms supported by these commitments. The alternative is to give up our particular commitments; however, when we understand the relevant commitments to make up our practical standpoint or core set of aims and preferences, that option cannot be taken easily, if at all.

Here, the Humean appears to run into trouble: if the justification of norms is built on the means-end principle, what justifies this principle? Hampton has argued that reasons to take the means to our ends are no less mysterious than reasons that are independent from our ends.[65] She emphasises that if you reject categorical reasons because they do not fit with naturalism, you cannot simply help yourself to instrumental reasons. The means-end principle can only determine which norms we ought to follow in the authoritative sense if it has normative authority itself. It needs to be a serious mistake to fail to obey the norms supported by your commitments. And it appears that it cannot be up to our own preferences whether we ought to do what our preferences require.[66] However, we had deemed this categorical form of authority to be non-natural and alien. Now it seems that the Humean is ultimately committed to the very thing they set out to avoid. Hampton understands supposedly 'natural' theories of normativity to fail on these grounds:

> using [the instrumental] conception of reason as a tool to naturalize other norms is just using another normative tool that is at least as 'unnatural' as the norms it is intended to naturalize. It therefore cannot serve as the Archimedean lever for the naturalization of other norms (...).[67]

A potential solution to Hampton's problem is offered by constructivist theories with a focus on constitutive rules of practical reason, like Street's: the means-end

[65] Hampton, *The Authority of Reason*, 125–66.
[66] Hampton, 140 n. 22.
[67] Hampton, 6.

principle has authority in virtue of being constitutive of our normative judgements.[68] To judge that we have a reason to φ, we must also judge that we have a reason to take the necessary means to φ-ing. This step is not optional; it is necessitated by the concept of taking something to be a reason. The constitutive role of the means-end principle makes it inescapable. At the same time, the authority of this principle remains natural – it is the result of our attitudes. The norms that promote our values have authority for us because, given our values, we must value these norms. This arguably gives our normative commitments the right kind of normative significance.[69]

A problem for this move is that it faces Enoch's 'schmagency' objection to constitutivist accounts of normativity.[70] Enoch argues that the fact that some requirement is constitutive of agency does not entail, on its own, that complying with it is practically necessary; we could opt for something close to agency – schmagency – which lacks this constitutive requirement. Enoch questions the normative significance of the term 'agent' and not 'schmagent' applying to you. We can extend this objection to Street as follows: why does it matter that the means-end principle is inescapable when you make normative judgements? We can avoid it by refraining from making normative judgements, in Street's sense, and make judgements of which the means-end principle is not constitutive instead. Street would agree that making normative judgements is itself not unavoidable – you have no reason to make normative judgements prior to making any. Therefore, we are not unavoidably bound to the means-end principle.

If the means-end principle cannot be vindicated by it being constitutive of normative judgements, Humean constructivism ends up being in the same boat as theories that explain authority in terms of desires, preferences or aims.[71] It is not plausible that

---

[68] It seems that this requires a narrow-scope reading of the means-end principle: if you value X, you ought to φ. If you do not value anything, it is not the case that you ought to do what promotes your values; the 'ought' is internal to your normative judgements.

[69] Street claims that the means-end principle is not the only constitutive principle of practical reason. If that is true, there could be multiple ways in which our values justify norms, which would further strengthen the idea that these values are normatively significant. However, as I discussed in the previous chapter, I doubt there are other principles that are plausibly constitutive of normative attitudes. An exception is the principle that you cannot simultaneously make normative judgements that contradict each other, but this does not seem to help to justify particular norms.

[70] Enoch, 'Agency, Shmagency'.

[71] Of course, there are reasons to question the constitutivist element of Humean constructivism beyond the 'schmagency' objection - we may doubt that there are special normative attitudes which

the means-end principle is constitutive of these attitudes: as Street points out, you can coherently desire an end – such as staying alive – without desiring its necessary means – such as having a painful operation.[72] A desire-based theory would therefore have to ascribe authority to the means-end principle on the basis of desires that support this principle. And since all desires are contingent on the Humean view, the authority of the means-end principle becomes contingent too: it is not necessarily the case that every agent ought to comply with the norms that form the means to their own ends. Likewise, if Street's constitutivist solution fails, she must say that the means-end principle is authoritative for an agent only if this agent either values the principle or is committed to valuing it by their other values. Now we face contingency not only when it comes to the content of our commitments, but also with respect to the justifying relationship between our commitments and norms. For those who do not care about taking the means to their ends, no norms are made authoritative by their ends, which means there are no norms at all they ought to comply with.

The Humean can respond to this problem in the same way as to the original problem of contingency: even if agents without a commitment to the means-end principle are possible, they would be extremely rare, as well as so alien that it is not a major problem that they produce counterintuitive results. Imagine a person, Dave, who does not care at all about taking the means to his ends. He desires to pass his exam and knows that studying the material is a necessary means to achieving that. It is not so odd for him not to have a desire to study the material. Yet, the disinterest we are imagining goes deeper than that – he does not care at all about doing what is necessary to satisfy his desire, regardless of what it takes. This makes it very unlikely that he will manage to pass his exam, despite genuinely wanting to. Presumably, Dave will constantly act on his strongest inclination with no regard to what is conducive to his aims. There is an obvious explanation for why we are normally committed to a general policy of taking the means to our ends: without this, we will be highly unsuccessful at getting what we want. There is something deeply odd, if not incoherent, about having desires but

are distinct from desires and beliefs, or that the means-end principle is constitutive of these attitudes.

[72] Sharon Street, 'Coming to Terms with Contingency: Humean Constructivism about Practical Reason', in *Constructivism in Practical Philosophy*, ed. James Lenman and Yonatan Shemmer (Oxford: Oxford University Press, 2012), 43.

not being committed to doing what it takes to satisfy them. It can be argued that we do not need to take a person with such an unnatural attitude towards their own commitments seriously in our theory of normativity. In this way, we can reject Hampton's claim that the authority of the means-end principle must be categorical in order for it to do its work in a theory of normativity. Apart from rare, possibly merely hypothetical exceptions, we do have commitments that make the means-end principle authoritative for us. This, in turn, gives a practical necessity to the norms our desires or normative judgements require us, through the means-end principle, to follow. The Humean then argues that this form of practical necessity counts as normative authority.

4. Conclusion

This chapter explored the implications and plausibility of Humean theories that explain the authority of norms by appealing to agents' contingent commitments. I illustrated how norms do or do not come out as authoritative for particular agents on such a view and highlighted the upshot that conventional norms can be authoritative, which I argued to be a desirable result. Next, I showed how Humeans can respond to several intuitively strong objections by either denying the problem or minimising it. In the end, we have to ask whether the disadvantages of a Humean understanding of normative authority – particularly its contingency – balance out against the problems it avoids, such as the costs of positing authority as a non-natural property. Philosophers on opposite sides of the meta-normative spectrum tend to have different views on the relative weight of the costs and merits of the theories on offer, and it is unclear how an objective verdict can be made here. I conclude, however, that it is far from obvious that a Humean explanation of authority is not correct.

Over the past three chapters, I argued that (1) the concept of authoritative normativity does not entail that it cannot be reduced to natural facts, but instead leaves room for commitment-based theories; (2) we can understand authoritative normativity to be grounded in justified norms, where our commitments are one possible source of norms' justification; and (3) the major substantive objections to contingent-commitment-based theories do not establish that this approach to normative authority fails. I

have not given an elaborate positive argument for Humean theories and have not defended a specific one. My conclusion is therefore not that a Humean theory is correct, but rather that what such a theory would establish is indeed authoritative normativity – the normative significance that our contingent commitments give to norms *can* count as the authority we are looking for. Humean theories offer a valid alternative to a non-naturalist understanding of normative authority, which demands serious and in-depth consideration. Although this result is more modest, it is nonetheless significant: it means that we do not need to choose between accepting that there are *sui generis* normative properties or accepting a normative error theory, because there is a serious third alternative. If irreducibly normative properties are too strange to accommodate, we can nonetheless save authoritative normativity by arguing that it arises out of metaphysically unproblematic commitments and norms. To eliminate this option, the error theorist must engage with the details of different Humean theories and come with additional substantive objections, without relying on a debatable non-naturalist interpretation of what authority is. We do not need to understand authority as something strange.

CHAPTER 5

The Conceptual Thesis of Moral Error Theory

0. Introduction

Having discussed how to avoid an error theory about normativity in general, I now turn to moral error theory. Morality is treated as an especially important system of norms. We often include moral considerations in our deliberations and tend to be motivated by them. We judge others' actions on their moral value, condemning those who violate moral requirements and praising those who take them to heart. However, according to moral error theory, we are mistaken to think that anything is morally right, wrong, required, permissible or impermissible at all. Moral error theorists deny the truth of all substantial moral claims. Here, 'substantial moral claims' excludes claims such as 'nothing is morally required' but includes all claims like 'saving a child within reach from drowning is morally required' and 'torturing someone for fun is morally wrong'. According to some moral error theorists, all such claims are false;[1] according to others, they are neither true nor false.[2] Since we normally do seem to

[1] Olson, *Moral Error Theory*.
[2] Joyce, *The Myth of Morality*.

intend to capture the truth with our moral judgements, something goes deeply wrong if they are never true. As I will discuss in the next chapter, error theorists need not say that we should discard our moral discourse and practice altogether; still, it is a significant loss to accept that they are based on falsehoods. While I have offered Humean theories as an antidote to normative error theory, it may be doubted that they can save us from a moral error theory – prominent moral error theorists present morality as necessarily involving an authority that is not based on agents' contingent commitments. With this in mind, a critical examination of moral error theory – especially its assumptions about morality's authority – is called for.

Commonly, moral error theorists deny that moral claims are true because they take moral properties to be unlike anything that exists. The thought is that these properties involve a mysterious feature which the world does not supply. Arguments for error theory involve two steps. The conceptual thesis claims that moral properties necessarily have a certain feature, as a matter of the meaning of our moral terms. The substantive thesis, in turn, denies that this feature exists or is instantiated. The two obvious (but not only) ways to defeat the conclusion that moral properties do not exist are to reject either the conceptual thesis or the substantive thesis. If we struggle to accommodate the feature that error theorists ascribe to moral properties, it is very important to examine whether the conceptual thesis is correct. In case it is a mistake to connect claims about what is morally wrong, right or required to the mysterious feature, the threat of scepticism about moral truth is averted.

In order to assess moral error theory, we therefore need to be clear on the conceptual thesis it is committed to. This is not as straightforward as one might expect. First of all, we cannot speak of a single unified theory; error theorists endorse various grounds for seeing moral properties as suspect. Secondly, even those error theorists that ultimately endorse the same thesis express it in different and rather abstract terms. To get a better grasp of the supposedly problematic feature of morality, I will examine the details of the conceptual thesis in the moral error theory of J.L. Mackie, Jonas Olson and Richard Joyce in turn. I will argue that the central claim of both Olson and Joyce is that morality necessarily has categorical normative authority. Then, I will turn to a critical assessment of this conceptual thesis. My criticism will build on the discussion of authoritative normativity of the previous chapters, as well as on the doubts about conceptual internalism I raised in chapter one. I will explain that the evidence

given for the categoricity of morality does not give the error theorists what they need; it is merely evidence that moral norms are categorical in the formal sense. The claim that morality would have to be categorical specifically in the authoritative sense, I will argue, is an implausibly narrow interpretation of the conceptual commitments of ordinary users of moral concepts. Without a credible conceptual thesis, moral error theory fails.

1. Mackie's conceptual thesis

To get a better understanding of the conceptual thesis that Joyce and Olson use, we should first look at what Mackie, their predecessor, identifies as the problematic element of our moral concepts. First, it should be noted that Mackie's defence of error theory relies on the same structure as more contemporary versions: he must identify something that is conceptually linked to morality in his conceptual thesis, and show this element to be fundamentally problematic in his substantive thesis. His most notable arguments, the argument from relativity and the argument from queerness, support his substantive thesis that the kinds of values that morality must consist of do not exist. Preceding this are claims that support his conceptual thesis: that moral values, if they exist, are indeed necessarily of this kind. If this cannot be upheld, the arguments for his substantive thesis lose their significance.

What is Mackie's conceptual thesis, then? There are different claims he makes about ordinary moral discourse. First of all, he rejects non-cognitivist analyses of moral language, arguing that it is a mistake to think that moral language is used to express non-cognitive attitudes and not to make assertions. Secondly, Mackie claims that moral assertions are about objective rather than subjective matters. When someone says that something is wrong, they are not making an assertion about their or their society's feelings about it, for example; they are asserting that it has a property of wrongness which is really out there, mind-independently. When Mackie calls values objective, he means that they are "part of the fabric of the world".[3] He infers that

[3] Mackie, *Ethics*, 24.

moral values are supposed to be objective in this sense from our ordinary use of moral language and the major traditions in Western moral philosophy.

However, the conceptual thesis of Mackie's error theory is not merely that moral values or properties are objective; in his view, the truly problematic feature of morality is its supposed practical authority. In his view, an objective good would have "to-be-pursuedness somehow built into it".[4] He believes that naturalist theories of morality cannot accommodate 'to-be-pursuedness'.[5] According to Mackie, this means that naturalist theories do not get moral concepts right: "Any analysis of the meanings of moral terms which omits this claim to objective, intrinsic, prescriptivity is to that extent incomplete".[6] In combination with the substantive thesis that this kind of prescriptivity does not exist, his error theory emerges: "a theory that although most people in making moral judgements implicitly claim, among other things, to be pointing to something objectively prescriptive, these claims are all false".[7] The question, then, is what Mackie means by prescriptivity. His discussion of this is unclear, because he does not clearly distinguish between motivating and normative force. In some passages his point appears to be that moral judgements must have the power to motivate agents to act in accordance with them, while some passages suggest that what he is referring to is normativity, and in others his conceptual thesis seems to contain a mix of both.

We could try to get more clear on Mackie's conceptual thesis by focusing on his defence of his accompanying substantive thesis; however, that still does not settle which interpretation is correct. Olson distinguishes four separate aspects of moral facts that Mackie uses a queerness argument against: supervenience of moral facts on natural facts, moral knowledge, moral motivation and irreducible normativity.[8] Joyce also pursues an interpretation of Mackie's argumentation in which the conceptual thesis of his error theory is about motivation. Joyce interprets it as a commitment to motivation internalism, which he defines as follows: "It is necessary and *a priori* that any agent who judges that one of his available actions is morally obligatory will have

[4] Mackie, 40.
[5] Mackie hereby shows a clear commitment to the common assumption that normative authority cannot be captured by metaphysically naturalist accounts, as discussed in chapter two.
[6] Mackie, *Ethics*, 35.
[7] Mackie, 35.
[8] Olson, *Moral Error Theory*, 79–138.

some (defeasible) motivation to perform that action."[9] Olson denies that Mackie is committed to motivation internalism, but does identify a motivational queerness argument in Mackie's work. In Olson's interpretation, the link that Mackie treats as conceptually necessary but problematic is between motivation and moral facts, instead of moral judgements: "Mackie thought that according to the ordinary concept of moral facts, such facts exert motivational pull on anyone acquainted with them."[10] The difference with motivation internalism is that this does not require false moral judgements to provide motivation. The accompanying claim is that this kind of intrinsic motivational pull is strange, which leads Mackie to conclude that it does not exist. If it is nonetheless true that the concept of morality requires it to exist, this concept is fundamentally flawed.

Both Joyce and Olson reject this motivation-based route to error theory. This version of the substantive thesis is not hard to defend: it does not seem to be the case that either moral judgements or knowledge of moral facts always bring some motivation.[11] However, regardless of the precise interpretation, this version of the conceptual thesis is weak: it is implausible that ordinary moral discourse supposes that either all moral judgements or all moral facts have an intrinsic power to motivate, so that any concept without this feature cannot be a moral concept. Joyce doubts that ordinary moral discourse comes with a commitment to motivation internalism.[12] According to Olson, our ordinary moral concepts do in fact leave space for agents not to be motivated by moral judgements or knowledge.[13] He also rejects the queerness arguments based on supervenience and moral knowledge.

What remains, as the inspiration for Olson's and Joyce's own argument for error theory, is Mackie's argument based on morality's supposed normativity. Olson makes clear that this is what he builds on.[14] Joyce also abandons the motivation-based argument for error theory in favour of arguing that a form of normativity which is in fact problematic is featured in our moral concepts. Their conceptual thesis, then, is about

---

[9] Joyce, *The Myth of Morality*, 18.
[10] Olson, *Moral Error Theory*, 109.
[11] See Joyce, *The Myth of Morality*, 17–23.
[12] Joyce, 19.
[13] Olson, *Moral Error Theory*, 112–15.
[14] Olson, 116–38.

a necessary link between morality and normativity. The important question is what kind of normativity this is. Here, again, Mackie can lead us in the right direction. What he sees as a feature of our moral concepts is not merely that there is always a pro tanto reason to act on moral requirements, but that this reason comes from those requirements themselves. In Mackie's terms, the prescriptivity is supposed to be intrinsic to moral values – ordinary moral judgements include a call for action that is absolute, rather than dependent on the desires of the agent. He calls this "the categorical quality of moral requirements".[15] He thus uses Kant's distinction between hypothetical and categorical imperatives to identify the necessary but ontologically problematic feature of moral properties: "my thesis that there are no objective values is specifically the denial that any such categorically imperative element is objectively valid."[16]

While this forms the basis of the moral error theory of Olson and Joyce as well, it is useful to note that there is a significant variety in their use of 'categorical'. Mackie defines a categorical imperative as expressing a reason for action that is "unconditional in the sense of not being contingent upon any present desire of the agent to whose satisfaction the recommended action would contribute as a means."[17] Therefore, what he takes to be fundamentally problematic about moral properties is that they would come with requirements that do not depend on desires. In the next section, I will show that there is an important distinction to be made in what it could mean for moral requirements to be categorical in this sense.

2. Categorical requirements and authority

The thesis that morality would have to involve categorical imperatives appeals to the widely accepted idea that moral requirements are in some important way not conditional upon the desires or interests of an agent. If, in order to see what the morally right action in a certain scenario is, someone merely considers which action promotes their own desires, they seem to be deeply mistaken about what morality is about and what counts as moral considerations. While the agent's interests may be morally

[15] Mackie, *Ethics*, 33.
[16] Mackie, 29.
[17] Mackie, 29.

relevant, 'Which action is morally right?' needs to be answered with reference to different aspects of the situation than 'Which action satisfies my desires best?'.[18] Michael Smith expresses this intuition as follows: "By all accounts, claims about what we are morally required to do are categorical imperatives: that is, if true, they are made true by facts about the circumstances in which we find ourselves; they cannot be defeated by the simple observation that acting in the way required will not serve any desire or interest that we happen to have."[19]

In "Morality as a System of Hypothetical Imperatives" Philippa Foot likewise argues that what sets categorical imperatives apart from hypothetical imperatives is that the first are not defeated by the information that an action is not a means to an end of the agent in question. And this is true of moral judgements: "the agent cannot rebut an assertion about what, morally speaking, he should do by showing that the action is not ancillary to his interests or desires."[20] If we tell someone to take the 10 a.m. train to Edinburgh, but they inform us that they no longer have an interest in being in Edinburgh, our prescription becomes invalid. However, if we tell them that it is wrong to beat up a fellow passenger, but they explain that they take great pleasure in committing violence and it would be in their best interest to do that, our judgement still stands. The application of moral judgements is unconditional upon a positive link to the agent's desires. Therefore, in Foot's view, moral judgements have a categorical quality.

Yet, the claim that morality is categorical can mean more than this. As Foot points out, many philosophers are not satisfied with the account of the categoricity of morality she accepts. The reason for this becomes apparent when she shows that categorical imperatives, specified in the way just explained, are certainly not exclusive to the moral realm: the validity of judgements of etiquette, too, are not conditional upon a relation to the agent's desires.[21] There is no retraction necessary when we tell someone to put their water glass to the left of a wine glass but they inform us that they enjoy having it on the right; it remains true that, in terms of etiquette, it is inappropriate to

[18] A moral egoist, who believes that what an agent morally ought to do is whatever is in that agent's best interest, would deny this.
[19] Smith, *The Moral Problem*, 84.
[20] Foot, 'Morality as a System of Hypothetical Imperatives', 1972, 307–8.
[21] Foot, 308–9.

do so. This shows that norms of etiquette are also categorical. The same goes for club rules, grammar, fashion norms or rules of a game. However, as Foot recognises, when philosophers talk about the categorical quality of morality, they usually mean to refer to something that sets it apart from institutions like etiquette and grammar. Mackie, Olson and Joyce certainly mean to. It is clear that when they characterise morality as consisting of or entailing categorical imperatives, they are ascribing something to it that etiquette is thought to lack. In Foot's words, what philosophers like Mackie, Olson, and Joyce have in mind when they conceptually connect morality to categorical imperatives is a "special dignity and necessity"[22].

The special dignity ascribed to morality brings us back to the distinction between formal and authoritative normativity. While formal normativity is a matter of norms determining which actions are correct or allowed and which are not, authoritative normativity concerns what we, in a significant sense, really have reason to do. As I have discussed, it is usually assumed that morality is authoritatively normative, while etiquette is brought forward as an example of norms with mere formal normativity.[23] Nonetheless, requirements of etiquette, like moral requirements, are inescapable in the sense that they do not become invalid if the recommended action is not in the agent's interest. No matter my preferences, putting my water glass to the right of my wine glass is incorrect. However, depending on the circumstances and my interests, I may have no authoritative reason to put my wine glass on the right. Requirements of etiquette are not inescapable in the sense that they unconditionally matter for what you should do. Therefore, despite its categorical application, etiquette does not have categorical authoritative normativity.

This example highlights that categoricity can relate to two separate aspects of norms: their application or their normative authority. Both formal and authoritative normativity can be categorical. When Foot claims that the requirements of morality are categorical in just the same way as the requirements of etiquette are, she is talking about formal normativity. Formally normative requirements can be categorical in the sense of making actions (in)correct for an agent regardless of that agent's desires. On

[22] Foot, 308.

[23] Although I argued against the assumption that etiquette is merely formally normative in chapter four, I will go along with it in this chapter. In this section, the important idea is that etiquette is not categorically authoritative, which I do not object to.

the other hand, the claim that moral requirements are categorical imperatives in a more special sense has to be the claim that they are categorically authoritative – that is, that they have a categorical form of authoritative normativity. On this interpretation, the idea is that everyone has authoritative reasons to do what is morally required, regardless of how the moral norms relate to the agent's interests.

3. Olson's conceptual thesis

Olson commits to this second claim in his conceptual thesis about morality. In his view, the conceptual thesis is "that moral facts are or entail facts about categorical reasons (and correspondingly that moral claims are or entail claims about categorical reasons)".[24] In combination with his substantive thesis that there are no categorical reasons, this leads to moral error theory. Initially, in his "In Defense of Moral Error Theory", he fleshes out this conceptual thesis as follows:

> To say that there are categorical reasons for some agent, A, to behave in some way, φ, is to say that there is reason for A to φ irrespective of whether A's φing would promote satisfaction or realization of some of A's desires or aims, or promote fulfilment of some role A occupies, or comply with the rule of some activity A is engaged in.[25]

It is striking that Olson hereby uses a definition of 'categorical' that is more demanding than how it is traditionally used, including by Mackie and Foot. Mackie contrasts 'categorical' with 'dependent on a present desire of the agent'.[26] Foot follows Kant in including a little more than this in the non-categorical category: an imperative is hypothetical rather than categorical not only if it relates to an agent's current desires, but also if it relates to their long-term projects or prudential interests.[27] Olson, however, says much more than this: for him, it is not just contingency upon a desire or interest

[24] Olson, 'In Defense of Moral Error Theory', 62.
[25] Olson, 64.
[26] Mackie, *Ethics*, 29.
[27] Foot, 'Morality as a System of Hypothetical Imperatives', 1972, 306–7.

of the agent that makes imperatives non-categorical; for something to be categorical, it must also not be conditional upon fulfilment of a role or compliance with a rule. His definition thus prevents us from using 'categorical' for formal normativity: we cannot say that a requirement subscripted to a rule applies categorically if that entails an independence of rules. With his understanding of categoricity, there is merely a non-categorical requirement on chess players not to move a rook diagonally, for example.[28] Olson could thus reject the claims Foot makes about what counts as categorical imperatives merely on the basis of his different, particularly demanding definition of 'categorical'.

After this first explanation of his conceptual thesis, Olson makes it even more clear that it focuses on authoritative normativity, not formal normativity. He rejects the view that there can be categorical reasons which do not actually count in favour of an action for an agent – in other words, which lack normative significance. He elaborates on his former characterisation of categorical reasons to reflect this: "to say that F is a categorical reason is to say that F matters normatively, irrespective of agents' desires, aims, or roles."[29] It is facts about categorical reasons thus understood that moral facts are or entail, according to Olson's conceptual thesis. His substantive thesis then denies there to be this kind of categorical reason. His terminology leaves no place for categoricity at the level of formal normativity at all. A rule may apply and make a certain action correct relative to it without this depending on an agent's interests or concerns, but that does not make it a categorical requirement in Olson's view. He only speaks of categoricity in the realm of authoritative reasons. And even there, he is especially demanding about what it takes for authoritative reasons to be categorical, as they must not only be independent of agents' interests and desires, but also independent of rules or roles.

Since we have seen how much clarification is still needed after claiming that moral facts are conceptually linked to categorical reasons, it is not surprising that Olson abandons this formulation of the conceptual thesis in his *Moral Error Theory*. Here, he does not talk about the problematic feature of morality in terms of categoricity, but in terms of irreducible normativity: "moral facts are queer in that they are or entail facts that count in favour of or require certain courses of behaviour, where the

[28] Olson, 'In Defense of Moral Error Theory', 65.
[29] Olson, 71.

favouring relation is irreducibly normative."[30] As I discussed in chapter two, Olson claims that his distinction between reducible and irreducible normativity corresponds to Parfit's distinction between the rule-involving and reason-involving senses of normativity and Hattiangadi's distinction between norm-relativity and normativity.[31] It is clear that these, in turn, are the same as what I call the distinction between formal and authoritative normativity. In Olson's view, a claim about an action being correct or incorrect according to some norm is reducibly normative.[32] It reduces to a claim about what a certain standard of correctness requires, which does not appeal to irreducibly normative reasons. By 'irreducibly normative reasons', in turn, Olson means facts that favour an action where the favouring relation "is not reducible, for example, to facts about what would promote satisfaction of [the agent's] desires, or to facts about [the agent's] roles or engagement in rule-governed activities."[33]

This shows that the difference Olson sees between reducible and irreducible normativity does not simply track the basic formal/authoritative divide. There is something similar going on here as in Olson's definition of 'categorical': he makes the category that he puts moral reasons in especially demanding by adding his own specifications that are not essential to that category. With 'categorical', he added that categorical reasons must be independent of rules and roles, not just independent of an agent's desires or interests as normally understood. Now, he also makes it true by definition that authoritatively normative reasons cannot be a matter of desire-satisfaction. Olson is free to use a term like 'irreducibly normative' for the particular kind of reason he has in mind; however, it is misleading to imply that this corresponds, in my terms, to 'authoritatively normative'. As I discussed in chapter two, Olson fills in the formal/authoritative distinction with a restrictive view on authoritative reasons, with the result that he, as a matter of definition, leaves no room on the authoritative side for types of reasons that arguably could be authoritative. He banishes many types of reasons, such as those based in our desires, to the formally normative side, even though the common-sensical idea seems to be that those reasons do matter. We can also conceive of reasons based in roles or rule-governed activities as authoritative, as long as those roles

[30] Olson, *Moral Error Theory*, 117–18.
[31] Olson, 121.
[32] Olson, 120.
[33] Olson, 122.

or activities have normative significance. Overall, Olson's irreducibly normative reasons are only a subset of what could fall under the authoritative in the fundamental division between formal and authoritative normativity.

The result is a rather narrow conceptual thesis. According to Olson, moral facts are or entail reasons that belong to the exclusive category of irreducibly normative reasons. This comes down to the claim that necessarily, if an action is morally required, you have an authoritative reason to do this action, without this depending on your desires, aims, roles, or engagement with rules and standards. However, Olson denies that such completely independent authoritative reasons exist, on the basis that they would be "metaphysically mysterious".[34] This is how he arrives at his denial of moral facts.

If Olson's conceptual thesis were simply that morality would have to be authoritatively normative, its accompanying substantive thesis – that nothing is authoritatively normative – would be undermined by my discussion of authoritative normativity earlier in this thesis. Authoritative normativity is not necessarily non-natural; it could be grounded in natural facts about agents. When we reveal in this way that authoritative normativity need not be mysterious, we demystify morality at the same time. Even if moral concepts come with a connection to authoritative normativity, this does not mean that moral properties are too strange to exist, since authoritative normativity need not be too strange to exist. The error theorist can try to argue against metaphysically naturalist theories of authoritative normativity, but since there is conceptual room for them, and since the obvious objections to them are not conclusive (as argued in chapter four), it is at least far from clear that the substantive thesis 'Nothing is authoritatively normative' is true. This argument for moral error theory, then, would not be convincing.

Olson can instead stick to his conceptual thesis that morality is authoritatively normative in a particular narrow sense. The Humean theory of authoritative normativity I have explored makes the authority of morality dependent on the commitments of the agent in question. It can therefore not establish that morality has necessary or categorical authority. If the error theorist ascribes specifically this type of authority to morality in their argument, the Humean will have to accept the substantive thesis and

[34] Olson, 136.

target the conceptual thesis instead. After clarifying the understanding of morality behind Joyce's error theory, I will assess both his and Olson's conceptual thesis together.

4. Joyce's conceptual thesis

Joyce's conceptual thesis, like Mackie's and Olson's, links morality to categoricity. Joyce believes that use of moral language comes with a commitment to the existence of categorical imperatives. Joyce uses 'categorical' in a more traditional way than Olson – for Joyce, a categorical imperative is one whose legitimacy does not depend on an end or desire of the relevant agent.[35] He often expresses his point by ascribing 'inescapability' to morality, arguing that, as a matter of meaning, moral requirements cannot be escaped by having no desires or ends that are satisfied by doing the prescribed act: "When we morally condemn a criminal, we do so with a force that implies 'regardless of whether it suits you.'"[36] Whereas, by ascribing categoricity to morality, Olson instantly targets desire-based, role-based and rule-based accounts of moral normativity, Joyce only focuses on denying that moral requirements are conditional upon the agent's desires and interests.

Like Olson, Joyce rejects an account of morality as having the same categoricity as etiquette. Joyce does this by making a distinction between weak and strong categorical imperatives and claiming that moral requirements are of the strong kind. Weak categorical imperatives apply and determine what is correct to do, regardless of whether they have any hold over the agent. They are categorical imperatives because the agent's desires are irrelevant. What separates them from the strong sort is that weak categorical imperatives "need not *bind* [the agent], they need not be *his* rules; they do not present *the thing to do*; he may legitimately ignore them."[37] If rules do apply to a person, what does it mean that they are not *his* rules? The best interpretation of Joyce's claim here is that these rules are not authoritative for this person. Joyce's distinction between weak and strong categorical imperatives maps onto the formal/authoritative

[35] Joyce, *The Myth of Morality*, 35.
[36] Joyce, 32.
[37] Joyce, 36.

distinction. What Joyce sees as strong categorical imperatives is what Foot would call categorical in the 'dignified' sense.

Crucially, Joyce claims that moral prescriptions are strong categorical imperatives – thus, like Olson, he puts them in the authoritative category. According to Joyce, we would have to fundamentally misunderstand the concept of moral requirements to say that while someone ought not to kill according to the rules of morality, what she really ought to do is to kill.[38] As a matter of the meaning of moral terms, he claims, moral requirements are authoritatively normative. And, as is very significant for his argument, this is not just any kind of authoritative normativity: our reason to comply with moral requirements must be categorical. Joyce believes that "it is part of our moral conceptual framework" that morality "binds everyone regardless of their ends".[39] In this view, it is not a contingent matter whether one has an authoritative reason to do what is morally right; morality is "imbued with an 'automatic reason-giving force'".[40] Joyce starts his argument for error theory with the claim "If *x* morally ought to ϕ, then *x* [morally] ought to ϕ regardless of whether he cares to, regardless of whether ϕing satisfies any of his desires or furthers his interests."[41] This first premise leads to the claim that "if *x* morally ought to φ, then *x* has a reason for φing regardless of whether φing serves his desires or furthers his interests",[42] where this reason must be understood as an authoritative one. His substantive thesis is then that there are no authoritative reasons that are independent of agents' desires and interests.[43]

---

[38] Joyce, 36–37.

[39] Joyce, 62.

[40] Richard Joyce, 'Reply to "On the Validity of a Simple Argument for Moral Error Theory"', *International Journal of Philosophical Studies* 24, no. 4 (2016): 519.

[41] Joyce, 518. This is Joyce's slightly updated version of the premise in *The Myth of Morality,* p. 42, to correct the error discussed in Kasper Højbjerg Christensen, 'On The Validity of a Simple Argument for Moral Error Theory', *International Journal of Philosophical Studies* 24, no. 4 (2016): 508–17.

[42] Joyce, *The Myth of Morality*, 42.

[43] The fundamental thought behind Joyce's conceptual thesis is that morality is inescapable. He might accept that morality's authority could depend on desires as long as everyone's desires are guaranteed to provide them with reasons to be moral. However, he argues that this is not the case: there is no inescapable desire that justifies morality for everyone, regardless of their circumstances, nor will everyone necessarily have some contingent desire that is satisfied by being moral (Richard Joyce, 'The Fugitive Thought', *The Journal of Value Inquiry* 34, no. 4 (2000): 469–71). The dismissal of this desire-based option lies in the background of Joyce's conceptual thesis.

Although both Joyce and Olson connect morality to categorical authority in their conceptual thesis, the views of normativity their error theories are based on are very different, as becomes clear from how they argue for their substantive thesis. As discussed, Olson believes that moral requirements are problematic because they would have to be irreducibly normative, which is a mysterious property he does not accept. Any less mysterious reasons for action that we might see as authoritative, such as those reducible to rules or desires, are banished to the category of mere formal normativity. This is the purpose of the qualification 'categorical' in his first formulation of the conceptual thesis, and the purpose of saying that morality's normativity must be 'irreducible' in his second formulation. His moral error theory follows from an error theory about authoritative normativity in general.

This is very different in Joyce's theory: he is not sceptical about all authoritative normativity, but only about the categorical kind. This is because he accepts an instrumentalist account of reasons, where reasons for action are always ultimately related to an agent's desires or ends. According to Joyce, an agent has reasons to do those actions that a fully informed and flawlessly deliberating version of that agent would advise them to do.[44] It is important that Joyce takes the advice of this idealised version of the agent to still depend on the actual desires of the agent - that is what the recommendations are ultimately based on. Hence, while Olson discounts reasons stemming from desire-satisfaction as merely formal, Joyce does take these, and only these, to be normative in the authoritative sense. In Joyce's picture, it is possible for authoritative reasons to be related to a system of norms, but only because the satisfaction of an agent's desires might depend on following those norms. We can now see the significance of him tying morality to categorical reasons: if all authoritatively normative reasons are based on desires, as Joyce believes, the reasons that are claimed to be essential to moral facts do not exist. Joyce must show that morality's authoritative normativity is of this specific desire-independent kind, or his error theory does not go through.

If we only see categorical requirements in the formal sense as essential to our moral concepts, an error theory would not get off the ground. Norms are not strange – we can explain in natural terms how a system of norms has come about, and once it is in place, we can explain in natural terms what it means for an action to be incorrect

[44] Joyce, *The Myth of Morality*, 75.

relative to that system. Formal normativity only requires the existence of social institutions, and those fit into a naturalistic worldview just fine. Indeed, Joyce is happy to grant that "nobody could doubt the existence of systems of rules", so that weak categorical imperatives can exist, since they merely rely on such systems.[45] Olson also concedes that non-categorical reasons, which I translated as reasons in the formal sense, are not strange.[46] Joyce admits that there would be no problem with morality and the notion of moral reasons if we thought of it as an institution which requires certain behaviour but not in an authoritative sense.[47] Therefore, if our moral concepts allow for such an understanding of moral requirements, we can avoid moral error theory. Clearly, a lot depends on how plausible the conceptual thesis of the error theorists is.

5. Rejecting the conceptual thesis

Both Joyce and Olson argue for a moral error theory by ascribing categorical authority to morality. Although Olson adds the specification that this authority must not depend on the agent's roles or rule-governed activities, the point of this seems to be to rule out that morality could be merely formally normative, which Joyce agrees with. I will therefore assess the conceptual thesis of Joyce and Olson as one.[48]

5.1 Categoricity

In defending the conceptual thesis, Joyce largely focuses on the claim that morality's demands are categorical. He emphasises that we do not make our moral judgements dependent on the desires of the agent in question: "The moral felon who convinces us that he desired to commit his crimes, that he had no desires that the actions thwarted, does not incline us to withdraw our judgment that he did what he ought not to have done. We do not permit him to evade his moral culpability by citing

[45] Joyce, 43.
[46] Olson, 'In Defense of Moral Error Theory', 65.
[47] Joyce, *The Myth of Morality*, 44.
[48] Finlay offers an alternative argument against Joyce's conceptual thesis in Stephen Finlay, 'The Error in the Error Theory', *Australasian Journal of Philosophy* 86, no. 3 (2008): 347–69.

unusual desires or interests.[49]" Joyce makes his case with the help of extreme examples, where a recurring one is our moral condemnation of the mass-murdering Nazi. Our condemnation is not the result of considerations about how the agent would have best satisfied their desires; the judgement that the Nazis morally ought not have committed mass-murder is distinct from the judgement that it was in their best interest to refrain from doing so, and the truth of the moral claim does not depend on the truth of the latter. Hurting others is wrong due to the harm inflicted on them, not due to the self-harm involved (if any).[50] To someone who argues that moral requirements bind us because it is in one's own best interest to respect them, Joyce raises the question: "Is that really why the Nazis were morally wrong in building their death-camps? – that in doing so they harmed themselves?!" He also raises the point that it would not make sense to punish these Nazis, thereby harming them, if all we accuse them of is doing something that harmed themselves. Overall, it is implausible that moral requirements are dependent on one's desires.

However, all these points merely show that morality is categorical in the formal sense, which is not the crucial claim under debate. As Foot showed, that a judgement is not made invalid by the agent lacking appropriate desires is just a sign of categoricity at the level of formal normativity. Opponents of error theory can grant that the content of moral norms and who they apply to do not depend on agents' desires, while denying that moral norms have a desire-independent authority. Harming others is not morally wrong because it constitutes self-harm – moral requirements are (mainly) fixed by different factors. This does not exclude the possibility that one's authoritative reason to refrain from doing what is morally wrong comes from the fact that it constitutes self-harm. Russ Shafer-Landau responds to Joyce's conceptual thesis by distinguishing two claims:

> (A) Morality is categorically applicable: a person can be morally obligated to do something even if doing it fails to serve any of her interests, wants, or needs.

[49] Joyce, 'The Fugitive Thought', 463.
[50] Joyce, *The Evolution of Morality*, 60.

> (B) Morality is categorically reason-giving: a person has a reason to do her moral duty even if doing her duty fails to serve any of her interests, wants, or needs.[51]

We should not mistake the plausibility of (A) for (B) being plausible as well.

Take another point Joyce makes to capture the inescapability he ascribes to morality: "Even the person who has rejected that whole realm we still think of as being under the jurisdiction of morality."[52] Foot accepts this: "There is, of course, a sense in which morality is inescapable. Consider, for instance, moral epithets such as 'dishonest', 'unjust', 'uncharitable'; these do not cease to apply to a man because he is indifferent to the ends of morality: they may indeed apply to him because of his indifference."[53] The reason she can accept this is that it merely relates to the formal normativity of morality: moral norms apply to us. It may seem special that one cannot escape their application – Foot points out that whereas you can escape being a good or bad spouse by not marrying, you cannot turn to a way of life in which you will not be counted as either morally good or bad.[54] However, we also cannot avoid being subject to the norms of etiquette or fashion, for example. Our way of life may only change which particular norms of etiquette or fashion apply to us. Finally, that moral blame is inescapable also does not establish anything beyond categorical formal normativity; as I argued in chapter four, it is appropriate to use norm-related criticism as long as the norm applies, regardless of its authority. We can condemn someone who burps as impolite and need not take this back if this person did not violate any of their own commitments by doing so.

Joyce tries to step from morality's categorical application to categorical authority with what he calls 'Mackie's Platitude': "It is necessary and *a priori* that, for any agent x, if x ought to φ, then x has a reason to φ."[55] The thought here is that a categorical ought specifically entails a categorical reason.[56] Shafer-Landau denies that this

---

[51] Russ Shafer-Landau, 'Error Theory and the Possibility of Normative Ethics', *Philosophical Issues* 15, no. 1 (2005): 109. In (B), 'reason' needs to be read in the authoritative sense.
[52] Joyce, *The Myth of Morality*, 31.
[53] Philippa Foot, 'Morality as a System of Hypothetical Imperatives', 2002, 167, n. 15.
[54] Foot, 167, n. 15.
[55] Joyce, *The Myth of Morality*, 38.
[56] Shafer-Landau, 'Error Theory and the Possibility of Normative Ethics', 110.

supposed platitude is indeed obviously true: we can ask whether we have a reason to perform our duty without engaging "in verbal trickery".[57] If we recognise mere formally normative reasons, however, we can accept a version of it: if x ought to φ, then x has a *formal* reason to φ. A categorical ought also entails a categorical formal reason. Yet, this keeps us in the realm of categorical formal normativity and does not tell us anything about the categorical authority of morality. The only interpretation of Joyce's claim on which it is a true platitude does not get him what he needs.

5.2 Authority

Both Joyce and Olson need to show that authority is a non-negotiable element of moral concepts, so that something cannot be a moral norm if it lacks this. Ordinary speakers – meaning, presumably, non-philosophers – need to imbue moral language with a commitment to morality being authoritatively normative. With this in mind, what would count as evidence for the conceptual thesis? If the error theorist is right about authority being inherent to moral concepts, the ordinary speaker should be disposed to or willing to combine moral judgements with claims about authority. It should also commonly feature in people's own ideas on moral matters that moral requirements are authoritatively normative. Of course, we cannot claim that terms like 'authoritative normativity' have a place in most people's thoughts; the very few people with this kind of vocabulary do not count as ordinary moral speakers. However, Joyce suggests that ordinary speakers do ascribe categorical authority to morality, just through more accessible language. Hence, we have to consider the possibility that everyday moral language reveals a commitment to categorical authority in a more implicit way.

There is one common element of moral discourse that may be thought to confirm that the ordinary speaker talks of moral requirements as authoritative. This potential piece of evidence is the prevalent use of the normative terms 'reason' and 'ought' in moral contexts. Ordinarily, if someone judges that it is wrong for A to φ, they will be disposed to agree that A ought not φ. Ought-claims are often used in moral contexts as if they do not add anything. Take the moral judgement that what the Nazis did was

[57] Shafer-Landau, 110.

wrong – if you accept this, it is odd, perhaps even worrying, if you ask whether the Nazis ought not have done what they did. A commitment to the claim that they ought not have done it seems to follow from a commitment to the moral judgement. Denying the ought-claim about the Nazis is worrying because it suggests you also deny the moral condemnation of them. Claims about reasons also have this connection to moral judgements: we tend to think that if some action is morally required, the agent has a reason to perform it. This general practice of thinking and talking about moral matters partly in terms of what we ought to do or have reason to do appears to be the best, if not only, available evidence for the normative element in the error theorist's conceptual thesis.

The problem for the error theorist is that 'ought' and 'reason' are not exclusively authoritatively normative terms; as I have argued, they have a formal use too. I have a formal reason to φ as soon as some norm exists that determines φ-ing to be correct. In contrast, I only have a reason in the authoritative sense to φ if the norm requiring it is authoritative for me.[58] Both are ordinary and valid usages. Joyce recognises this, as he also discusses reasons in relation to requirements that are merely formally normative. In the context of institutions that he does not see as authoritative, like etiquette, Joyce uses the term 'institutional reasons'. He contrasts these with *real* reasons which bind an agent and cannot legitimately be ignored.[59] In his view, "[r]easons are contained implicitly in the rules just in virtue of the presence of 'ought' statements."[60] He asserts that there is nothing linguistically mistaken about using terms like 'ought' and 'reason' to endorse a system of norms.[61] And when you do not endorse the norms in question, it is correct to use a reasons-claim with 'According to...' added to it – for example, 'According to etiquette, you have a reason to put your water glass on the left'. In this way, normative terms can also be used when referring to mere weak categorical imperatives – that is, requirements that are categorical in the formal sense only. This makes it difficult for the error theorists to prove their conceptual thesis. In Parfit's words: "when we ask whether some claim is normative in the reason-implying [i.e.

[58] David O. Brink, 'A Puzzle About the Rational Authority of Morality', *Philosophical Perspectives* 6 (1992): 8; Brink, *Moral Realism and the Foundations of Ethics*, 60–61.
[59] Joyce, *The Myth of Morality*, 40–41, 49.
[60] Joyce, 40.
[61] Joyce, 40.

authoritative] sense, it is not enough to appeal to the words with which someone states this claim. Normative words can be used in different senses."[62]

It would be a mistake to think that, in order to oppose error theory's conceptual thesis, we must argue that all normative-sounding claims in moral language are to be understood in a formal sense. This may be just as implausible as the thought that they are all meant in the authoritative sense. There is a significant middle ground between these positions, which still defeats the error theorist's conceptual thesis. I suggested in chapter one that we should not assume that ordinary users of moral concepts have specifically either formal or authoritative normativity in mind. The ordinary user of moral language is not a metaethicist. The distinction between formal and authoritative normativity is a technical one that even many philosophers do not recognise. It seems odd to say that everyday moral thought is imbued with a commitment to the authoritative form of normativity: ordinary speakers do not have a particular form of normativity in mind when they claim there is a moral reason to do something, and there are different theoretical interpretations possible of their language and behaviour. The meaning of normative claims linked to moral judgements could be indeterminate, like most people's conception of morality is, arguably, largely indeterminate. In the end, it is not the case that our moral language is at its core imbued with claims that undeniably point to authoritative normativity.

5.3 Categorical authority

Although we can question that there is any link to authority at all enclosed in moral concepts, we need not go this far – as I showed in chapter four, even Humeans can say that, in practice, morality is almost always authoritative. On a theory of authority focused on agents' contingent commitments, nearly everyone has authoritative reasons to do what is morally required, the only exceptions being persons with highly unusual desires or values. This captures the intuition that morality is especially important. Since all the Humean picture fails to capture is necessary or categorical authority, this is what the conceptual thesis has to focus on. Simon Kirchin points out a tension in moral error theories: error theorists need to find a conceptual commitment

[62] Parfit, *On What Matters: Volume Two*, 309.

in everyday moral thought and language which is plausibly mistaken, but which is still clearly a commitment.[63] For the substantive thesis to be plausible, the conceptual thesis needs to be highly specific, which – at least when the added details concern something as abstract as authoritative normativity – makes it less plausible itself. We can ascribe some general commitments to ordinary moral thought. However, Kirchin argues, we can reject the specific interpretation an error theorist must make of such a commitment to condemn it as mistaken. I take this to apply to the conceptual thesis of Olson and Joyce as well: ordinary speakers may have a commitment to morality being important or especially binding, but the interpretation of this vague intuition as 'morality has categorical authority' is too specific.

One might think that the requirement of categoricity does not add anything to the claim that morality has authority; normativity cannot be authoritative if it is not categorical (or necessary). I have argued against such overly limiting preconceptions about authority in previous chapters. A theory that treats authority as commitment-dependent and contingent has counterintuitive implications, but these can largely be minimised and do not disqualify such a theory. Joyce claims that accepting that people who genuinely don't care about morality are free to ignore it comes down to acknowledging that "the authority of morality is an illusion".[64] This reveals a particular understanding of authority, which we can reject – morality can have genuine authority for people even if it lacks authority for a few.[65]

Remember that the conceptual thesis is not that ordinary speakers normally – or even always – conceive of morality as having categorical authority, but that this is a non-negotiable element of moral concepts. The problematic commitment error theorist point to needs to be so deeply entrenched in moral thought that a person who does not commit to it is not thinking morally.[66] And, crucially, the truth of this commitment has to form part of the truth-conditions of moral claims. Terence Cuneo highlights a distinction between our conception and concept of morality: we can

[63] Simon Kirchin, 'A Tension in the Moral Error Theory', in *A World Without Values: Essays on John Mackie's Moral Error Theory*, ed. Richard Joyce and Simon Kirchin (Dordrecht: Springer, 2010), 167–82.
[64] Joyce, *The Evolution of Morality*, 63.
[65] The Humean can also respond that morality can be authoritative even for those who do not directly care about it – see section one of chapter four.
[66] Kirchin, 'A Tension in the Moral Error Theory', 169.

accept that our common-sensical conception of moral facts includes a commitment to categorical authority without accepting that categorical authority belongs to the *nature* of moral concepts.[67] This resembles the contrast I made in chapter one between conceptual internalism, which the error theorists argue from, and sociological internalism, the thesis that it is generally believed that morality comes with authoritative reasons.[68] We can give an explanation for such a belief: there is social pressure to learn to take morality seriously regardless of our own desires, because this is beneficial to society.[69] This does not mean that it is essential to moral concepts.

Consider two standard beliefs about witches: (1) they fly on broomsticks; (2) they perform magic. If we learn that no one can fly on broomsticks, we have made a mistake about what witches do. Upon learning this, we should adjust our conception of witches. However, if we learn that no one can perform magic, this has more serious consequences: since it is an essential part of the concept of a witch that she can perform magic, we should stop believing in the existence of witches upon learning that (2) is an error. The error theorist needs to show that ordinary speakers have a commitment to categorical authority that has the role of (2), not (1), in the moral case. It is far from obvious that we can ascribe a *conceptual* commitment to categorical authority to ordinary speakers; the error, if there is one, need not be so fundamental that all moral claims thereby become unsound.

If it is conceptually true that morality has categorical authority, all philosophers debating this question must be either confused or concerned with something other than morality. This is a highly uncharitable interpretation of a traditional and ancient subject. Whether we – universally, necessarily – have reasons to do what is morally required is an open question which requires more than an analysis of moral concepts. Those who deny that we have authoritative reasons for fulfilling our moral duty do not thereby change the subject; we recognise Foot as discussing morality even when she suggests it has nothing more than categorical formal normativity.[70] Shafer-Landau writes, "[t]he puzzlement that competent speakers sometimes feel when confronted

[67] Cuneo, 'Destabilizing the Error Theory', 90.
[68] Goldsworthy, 'Externalism, Internalism and Moral Scepticism', 41.
[69] Philippa Foot, 'Reasons for Action and Desires', in *Virtues and Vices and Other Essays in Moral Philosophy* (Oxford: Oxford University Press, 2002), 153.
[70] Shafer-Landau, 'Error Theory and the Possibility of Normative Ethics', 111.

with what they take to be their duty, and the many philosophical efforts to address the threat of amoralism, are, together, excellent evidence that Mackie's platitude [interpreted in the authoritative sense] is no platitude at all."[71] The serious philosophical interest in 'why be moral?' shows that the authority of morality is not self-evident.

Joyce may object that his conceptual thesis does not have the implausible amount of detail I have ascribed to it. Especially in his work after *The Myth of Morality*, he emphasises that the problematic feature of morality is very difficult, if not impossible, to capture in a precise way. In his eyes, we have a rough commitment to morality being "imbued with a deeply mysterious kind of force—a kind of primitive feeling of 'being bound by rules and ends' that resists explication".[72] Joyce calls this the 'fugitive thought'.[73] He believes that, as important as this thought is, we may not be able to clarify exactly what it comes down to: "Perhaps Mackie and I fumble to dissect something that by its very nature cannot be brought into the light to be picked over by philosophical scrutiny."[74] He suggests that this deep unclarity of the special feature we ascribe to morality even reinforces the error theorist's claim that it is problematic: "Indeed, it may be the very perniciously vague, equivocal, quasi-mystical, and/or ineliminably metaphorical imponderabilia of moral discourse that troubles the error theorist."[75] Joyce could therefore say that the claim that morality has categorical authority is only one particular attempt at capturing the problematic commitment in moral thinking; objections to that specific conceptual thesis do not disprove that there is indeed a more vague conceptual commitment to what he calls "a kind of mystical practical authority".[76]

Note that this appeal to vagueness makes the argument for error theory vague as well. Joyce suggests that there necessarily is more to morality than what the natural world can supply, but that what this extra element is cannot be formulated. How, then, could we become certain that the error theory is true? How could it be clear both

---

[71] Shafer-Landau, 111.
[72] Richard Joyce, 'The Error In "The Error In The Error Theory"', *Australasian Journal of Philosophy* 89, no. 3 (2011): 524–25.
[73] Joyce, 529.
[74] Joyce, 525.
[75] Richard Joyce, 'Morality, Schmorality', in *Morality and Self-Interest*, ed. Paul Bloomfield (Oxford: Oxford University Press, 2007), 61.
[76] Joyce, 60.

that the special feature forms a non-negotiable element of moral concepts and that it does not exist, when it is impossible to say what the special feature is? A fugitive thought is a weak starting point for an argument with a highly radical conclusion.

Moreover, with this line of thought, Joyce encounters the problem Kirchin pointed out for moral error theories in general: the conceptual thesis needs to be specific enough to make the substantive thesis plausible.[77] Joyce can keep the conceptual thesis vague to avoid the objections to it I raised, but this invites serious objections to the accompanying substantive thesis instead. Does the world supply a 'mystical practical authority'? Is the feeling that we are bound by rules and ends correct? That depends on what we mean by this. There are interpretations of these thoughts available that do not presuppose the existence of a strange, *sui generis* normative force. The Humean can say that morality is normatively special because it is authoritative for an overwhelming majority of persons: I have argued that anyone with commitments in the normal human range has authoritative reasons to be moral, whether they directly care about being moral or not. This gives morality an importance that we can capture by ascribing practical authority or bindingness to it. Finlay likewise argues that an end-relational theory of normativity can still account for the apparent difference between morality and other systems of norms:

> Moral standards or ends are of pressing concern to ordinary, decent people, and their importance to us typically overrides the importance of other standards and ends. This is sufficient to explain why we are much more serious and intransigent about our moral appraisals than we are about our appraisals of manners or chess strategy (...).[78]

Joyce cannot allow that these alternative interpretations of the special feature of morality are accurate; to maintain that the special feature is not instantiated, he needs to give a specific reading of it that excludes them. His preferred specific reading, that morality has categorical authority, is far less plausibly a conceptual commitment than the general thought that morality has a special practical importance.

---

[77] Kirchin, 'A Tension in the Moral Error Theory'.
[78] Finlay, 'The Error in the Error Theory', 354.

In response, Joyce may argue that our intuitions about particular cases are not in line with the Humean take on the special force of morality. He uses Plato's example of Gyges, a simple shepherd who finds a ring of invisibility and uses this to kill the king, undetected, and seize power.[79] Gyges acts immorally to his own benefit, with no risk of getting caught and facing punishment. Joyce imagines that Gyges does not harm himself in any way with this behaviour. What he does is nonetheless morally wrong. Moreover, Joyce argues, we treat morality as being authoritative even for Gyges: our condemnation of him goes further than the thought "[o]f course by killing an innocent person the shepherd is breaking a rule of morality, and so *according-to-the-rules-of-morality* he ought not do it".[80] We would say that Gyges has a "real reason" not to kill – "a reason that somehow engages the shepherd – one that he cannot legitimately ignore."[81] By stipulation, this authoritative reason cannot come from Gyges' own desires or values. This intuition is therefore not compatible with the Humean theory that morality has a desire-dependent authority.

We can deal with this example and others like it in the same way Street counters cases with ideally coherent eccentrics, as discussed in chapter four. Firstly, these examples are highly artificial and unlikely to occur in reality. This is especially clear in the case of Gyges: we must not only ascribe purely selfish desires and a lack of concern for others to him, but also the power of invisibility. Without these, he would plausibly have desires that are frustrated by immoral actions. Real shepherds would therefore have an authoritative reason not to kill, even on a Humean view. Secondly, Street points out that these kinds of examples are so alien that our intuitions about them are not reliable.[82] Our moral theories and intuitions are directed at real human beings whose commitments and circumstances are reasonably like ours. We may believe that even Gyges has authoritative reasons to be moral because we mistake him for one of us; if we would completely grasp his peculiarly selfish desires, and fully understand that he has no risk of facing negative consequences himself, we may not be so certain that he is bound by morality. Our everyday conception of morality, including the idea that it has a special practical authority, arises in and is sustained by a context

[79] Joyce, *The Myth of Morality*, 32–33.
[80] Joyce, 36.
[81] Joyce, 41.
[82] Street, 'In Defense of Future Tuesday Indifference'.

containing normal human beings, with other-regarding concerns and no magic rings. We cannot assume that this conception includes verdicts about very unusual kinds of persons, and it is especially unclear that our *conceptual* commitments about morality take someone like Gyges into consideration. This would make such a commitment, once again, questionably detailed. Overall, it is not obvious that the vague idea that morality is normatively special must be interpreted to contain a claim to categorical authority.

My diagnosis of the conceptual thesis about morality that Joyce and Olson use is the same as that of the standard conception of authoritative normativity: it is overly restrictive, adding mysterious elements that do not belong in the bare concept. Just as there is conceptual room for a commitment-based theory of authoritative normativity, there is also conceptual room for morality's authority to depend on one's commitments. In Kirchin's view, the "Achilles' heel" of moral error theory is that it links the legitimacy of moral claims to everyday moral commitments.[83] He points out that everyday moral thinking is messy: people have different commitments which are not necessarily compatible. It is controversial what ultimately lies at the core of moral thought, if anything.[84] We cannot expect to find a clean, specific claim about moral normativity in the vague idea of morality we share. The conceptual thesis is therefore implausibly detailed. While 'φ-ing is morally required and I have a *formal* reason to φ' is a tautology, 'φ-ing is morally required and I have an *authoritative* reason to φ' is not. As David Brink argues, the answer to whether moral requirements are authoritative for us depends on substantial facts about what the content of our moral obligations is, what we are like as agents, and what the nature of our reasons for action is.[85] We do not know that there are categorically authoritative moral requirements as soon as we know that there are moral requirements. As a consequence, the error theorist's argument fails: claiming that nothing has categorical authority is not sufficient for establishing that there are no true moral claims.

[83] Kirchin, 'A Tension in the Moral Error Theory', 182.
[84] Kirchin, 182.
[85] David O. Brink, 'Moral Realism and the Sceptical Arguments from Disagreement and Queerness', *Australasian Journal of Philosophy* 62, no. 2 (1984): 111–25.

6. Conclusion

A paradigmatic form of moral error theory arises from doubts about morality's authority: moral properties are thought to be imbued with a mysterious normative force. If we are unable to accommodate such a force – perhaps because we conceive of it as non-natural – the radical conclusion that there are no moral properties comes into view. In the previous chapters, I argued that authoritative normativity need not be mysterious. Joyce and Olson, however, ascribe more than authoritative normativity to morality: their conceptual thesis is that morality would have to be authoritatively normative in a categorical way. In their view, the mysterious normative force that makes moral properties strange is an authority that is independent of agents' desires or interests. Olson adds that this force must also be independent of agents' roles or rules, which reinforces the idea that he is concerned with authoritative and not formal normativity. I have argued that this conceptual thesis is not plausible. The evidence offered for the categoricity of morality merely establishes that moral norms *apply* to agents regardless of their desires and interests. The use of normative terms in connection to moral claims does not reveal a commitment to authoritative normativity, because these terms also have a formal sense, and it is likely that ordinary speakers do not distinctly have the authoritative sense in mind. The claim that morality has categorical authority is specific, going beyond the general thought that morality has an important practical force. It is implausible that ordinary users of moral language have a commitment to categorical authority that goes sufficiently deep to form a non-negotiable part of moral concepts. Instead, we can question whether everyone has authoritative reasons to be moral without changing the subject. Moral properties do not stand or fall with the existence of categorical authority.

Chapter 6

The Second Revolution of Moral Fictionalism

0. Introduction

If we were to discover that moral error theory is correct after all, how should we move on? This 'now-what problem' has received growing attention from error theorists themselves,[1] with good reason. Error theory may be thought to imply that we should get rid of moral thinking.[2] However, our moral practices run deep in the fabric of society, and the way we behave towards each other would be drastically different without them. This threat of radical practical consequences puts a significant burden on error theory. It raises the stakes of the debate, making it even more important to question the assumptions behind error theory and resist accepting that moral properties would have to be mysterious. Richard Joyce recognises this incentive for trying to remove the threat of radical practical implications from error theory: "If a persuasive case were to be made that we could adopt the error-theoretic position and civilization

[1] See, for example, Richard Garner and Richard Joyce, eds., *The End of Morality: Taking Moral Abolitionism Seriously* (New York: Routledge, 2019).

[2] In this chapter, I use 'error theory' specifically for moral error theory, not broader error theories about normativity.

would not collapse – that life would go on as before, or even go on *better*! – then the opposition to the theory might diminish, or at the very least lose some of its determination".[3]

With this ambition, Joyce presents his revolutionary moral fictionalism (hereafter referred to as 'fictionalism'). He sees morality as a useful fiction, due to the special advantages he ascribes to moral thinking. Joyce recommends that we preserve these advantages after rejecting moral beliefs as mistaken by *pretending* that some moral propositions are true. The appeal of fictionalism is that it allows error theorists to continue to deliberate in moral terms yet avoid being committed to supposed falsehoods. In contrast to hermeneutic fictionalism, the claim of revolutionary fictionalism is not that we already merely pretend that some moral propositions are true, but that we *should* do this. Joyce suggests that everything will more or less stay the same when we make this switch from moral belief to make-belief. In this way, fictionalism is supposed to invalidate concerns about the practical impact of collectively accepting error theory.[4]

I will argue that fictionalism fails to meet this promise, because it is more revolutionary than has been assumed. The problem is that fictionalism does not merely require us to change our attitude towards morality from belief to pretence; it also requires us to change the *content* of morality. In other words, a fictionalist should endorse a different set of moral obligations than moral believers conventionally endorse.[5] This becomes clear once we recognise that Joyce's instrumental justification of moral practices is fundamentally similar to that of David Gauthier's contractarianism. I will argue that, just like the contractarian morality, the revolutionary moral fiction lacks a substantial range of conventionally recognised moral obligations. Although my discussion is focused on Joyce's argument for fictionalism, the implications it uncovers apply more generally: the change in morality's content is a challenge for anyone who

[3] Richard Joyce, 'Moral Fictionalism: How to Have Your Cake and Eat It Too', in *The End of Morality: Taking Moral Abolitionism Seriously*, ed. Richard Garner and Richard Joyce (New York: Routledge, 2019), 151.

[4] Joyce, *The Myth of Morality*, 231.

[5] My focus, like Joyce's, is on fictionalism as a collective strategy for a group of error theorists; hence, my conclusions may not extend to a scenario where an isolated error theorist decides to adopt fictionalism.

advocates becoming a fictionalist for the sake of instrumental value[6] – especially when, as is standard, this is combined with Joyce's prominent form of moral error theory. The fiction's revolutionary content is a significant drawback for both fictionalism and error theory. To absolve error theory from the burden of having radical practical implications, showing that error theorists should not abolish moral practices completely is not enough; it also needs to be shown that they should preserve a wide range of moral norms resembling what we currently endorse. This is where fictionalism fails. The goal of this chapter is not to establish that fictionalism is the wrong answer to the now-what problem or that error theory is false, but rather to demonstrate that the former cannot remove the unappealing revolutionary nature of the latter. This, in turn, strengthens the previous chapter's conclusion that the conceptual thesis of Joyce and Olson is not plausible enough to warrant accepting that there are no genuine moral requirements.

1. The rationale for revolutionary moral fictionalism

Joyce's moral error theory and fictionalism are both motivated by his view of reasons.[7] As I have explained, Joyce argues that moral obligations necessarily come with categorically authoritative reasons. At the same time, he argues that such reasons do not exist; our authoritative reasons are fully dependent on our desires or preferences. Together, these claims entail that moral obligations do not exist. In light of his argument for error theory, Joyce can only argue that error theorists should adopt fictionalism by pointing to instrumental reasons to do so. In his view, the question of what to do with discredited moral thinking is settled by a calculation of what serves our ends best: "when morality is removed from the picture, what is practically called for is a matter of a cost-benefit analysis, where the costs and benefits can be understood liberally as preference satisfactions".[8] Consequently, the only basis on which Joyce can recommend that we adopt fictionalism is its instrumental value in maximising our individual

[6] E.g. Daniel Nolan, Greg Restall, and Caroline West, 'Moral Fictionalism versus the Rest', *Australasian Journal of Philosophy* 83, no. 3 (2005): 307–30.
[7] Throughout this chapter, I use 'reason' in the authoritative sense.
[8] Joyce, *The Myth of Morality*, 177.

preference-satisfaction. Although advocates of fictionalism can make adjustments to the exact view of instrumental reasons used, they must understand reasons to ultimately be dependent on the agent's aims or preferences, or else Joyce's type of argument for error theory would be undermined.[9]

There are two prominent alternative responses to accepting error theory that Joyce needs to dismiss as less useful than fictionalism. The first is abolitionism, which requires error theorists to get rid of moral thinking and stop using moral language altogether.[10] While abolitionists take moralising to cause more harm than good, Joyce argues that moral thinking is too valuable to abandon. He ascribes a special role in deliberation to moral beliefs: when we think of an action as morally required, we attach a 'must-be-doneness' to it.[11] While we may fail to act on merely prudential considerations due to weakness of will, moral thinking "functions to bolster self-control against such practical irrationality".[12] Joyce thus takes moral beliefs to be advantageous, despite being false, in virtue of their special motivational power. Of course, this is only beneficial if moral beliefs motivate actions that are conducive to our preference-satisfaction. Joyce assumes that, generally, this is the case: morality is supposed to promote sincere cooperation, which is itself instrumentally valuable.[13] I will elaborate on the instrumental value of moral behaviour below.

The second main alternative to fictionalism is conservationism, the recommendation that error theorists continue having moral beliefs and making moral assertions in

---

[9] I base my discussion on the basic instrumentalist view of reasons that Joyce consistently uses in his defence of fictionalism. In his defence of error theory, he accepts an alternative instrumentalist view according to which an agent has a reason to do an action if and only if an idealised version of them, who is fully informed and deliberating flawlessly, would advise them to do the action (Joyce, 53–79). Joyce emphasises that, on this second view, an agent's reasons still fully depend on their contingent desires (Joyce, 80–105). This aspect of reasons being relative to individual and contingent ends is crucial for the implications I will draw for the fiction's content. Whether the agent ought to pursue their actual or idealised (but still contingent and subjective) ends is less relevant for my argument. My argument would therefore still go through if fictionalism were defended based on Joyce's alternative instrumentalist view. The same applies, arguably, to any other view of reasons that error theorists could coherently endorse.

[10] Richard Garner, 'Abolishing Morality', in *A World Without Values: Essays on John Mackie's Moral Error Theory*, ed. Richard Joyce and Simon Kirchin (Dordrecht: Springer, 2010), 217–33.

[11] Joyce, *The Myth of Morality*, 181.

[12] Joyce, 184.

[13] Joyce, 181.

everyday contexts, despite being disposed to believe in error theory in critical contexts, such as the seminar room.[14] Against this, Joyce argues that making ourselves believe moral propositions while recognising evidence that they are false will have detrimental effects, since it violates the instrumentally valuable policy of aiming for true beliefs.[15] He concludes that maintaining genuine moral beliefs will not maximise error theorists' preference-satisfaction.

However, Joyce does not want to give up on moral thinking as a whole: he recommends that we avoid the harms of false beliefs yet preserve the benefits of seeing the world through a moral lens by becoming fictionalists about moral discourse. In most contexts, a moral fictionalist sounds just like a moral believer: they will call certain actions right or wrong and employ moral terms in their deliberation. What distinguishes hem from a moral believer is that the fictionalist does not actually believe moral propositions, but only pretends that they are true. Likewise, they merely pretend to assert moral propositions – their moral utterances lack assertoric force, as if they were speaking as an actor in a play. The fictionalist's commitment to error theory will only be apparent in contexts where the status of morality is discussed, where they will be ready to admit that their engagement with morality is merely an elaborate case of make-belief. In Joyce's view, his fictionalism is revolutionary in the sense that it requires a change from moral belief to make-belief.

As long as some of the positive impact of genuine moral beliefs remains when they get turned into make-belief, fictionalism has an advantage over abolitionism. Assuming that cultivating moral beliefs known to be false is too harmful and that there are no other viable options, this would mean that error theorists should engage with morality as a fiction. For the sake of argument, I will accept Joyce's conclusion that fictionalism wins this cost-benefit analysis.[16] My focus is on what the implications are of error theorists choosing this policy for the sake of their own preference-satisfaction.

---

[14] Jonas Olson, 'Getting Real about Moral Fictionalism', in *Oxford Studies in Metaethics 6*, ed. Russ Shafer-Landau (Oxford: Oxford University Press, 2011), 181–204.

[15] Joyce, *The Myth of Morality*, 178–79. For criticism of this argument, see Olson, 'Getting Real about Moral Fictionalism', 193–95.

[16] For objections against Joyce's argument, see Terence Cuneo and Sean Christy, 'The Myth of Moral Fictionalism', in *New Waves in Metaethics*, ed. Michael Brady (New York: Palgrave-Macmillan, 2011), 85–102; Björn Eriksson and Jonas Olson, 'Moral Practice after Error Theory: Negotiationism', in *The End of Morality: Taking Moral Abolitionism Seriously*, ed. Richard Garner and

After establishing that an error theorist should adopt an attitude of pretence towards morality, a substantial question remains: *which* moral propositions should they pretend to be true? Morality is not made up of a single, clear-cut set of propositions that any moral believer accepts. Since the contents of genuine moral beliefs differ, the content of the moral fiction can vary as well. As fictionalists, do we pretend that it is wrong to lie to the murderer at the door? Do we tell ourselves that we ought to donate a share of our income to those in need? There is an infinite range of specific moral fictions to choose from – which should a revolutionary fictionalist adopt? Importantly, since fictionalism is built on error theory, the answer cannot be that a fictionalist ought to endorse moral propositions that are true.

Instead, the appropriate criterion to assess moral fictions by is the same one that requires error theorists to adopt fictionalism rather than an alternative: the right option is the one with most instrumental value. The only reason an error theorist has to engage with morality as a fiction is that this is conducive to the satisfaction of their preferences. As a consequence, the specific moral fiction they should adopt is the fiction that it is most advantageous for them to engage with. A fictionalist should thus pretend that lying is wrong if and only if this is more beneficial to them than not doing so. The whole set of moral propositions that a fictionalist will pretend to be true will be determined in the same way. This follows directly from Joyce's argument for adopting fictionalism. Once error theory is accepted, the right version of morality is the most useful one. This has been recognised by Joyce,[17] as well as other commentators on fictionalism.[18] What has hardly been explored so far, however, is which moral fiction we can expect to be the most useful one.[19]

To be clear, the question here is which fiction we should adopt upon becoming fictionalists and continue to use thereafter. The idea is not that we should determine for each situation separately whether it is most advantageous to accept certain moral

Richard Joyce (New York: Routledge, 2019), 113–30; Matt Lutz, 'The "Now What" Problem for Error Theory', *Philosophical Studies* 171, no. 2 (2014): 351–71; Olson, 'Getting Real about Moral Fictionalism'.

[17] Joyce, 'Moral Fictionalism: How to Have Your Cake and Eat It Too', 154.

[18] Nolan, Restall, and West, 'Moral Fictionalism versus the Rest', 327; Olson, 'Getting Real about Moral Fictionalism', 189.

[19] Cf. François Jaquet, 'Utilitarianism for the Error Theorist', *The Journal of Ethics* 25, no. 1 (2021): 39–55.

propositions; this continuous adjustment would be cognitively demanding and motivationally ineffective. As Joyce describes it, the fictionalist does not choose to engage with morality only in those situations where moral motivation would help them, but adopts a constantly present "habit of ethical thinking".[20] For this habit to be effective, the content of one's moral thoughts should be constant as well. Thus, we are looking for a stable and continually present moral fiction that is overall the most advantageous to adopt.

Nolan, Restall, and West have suggested that which moral fiction is most useful will largely depend on current moral practices: "it will be easier to institute a fiction that is a close relation of moral theories currently employed, than to construct a new one out of whole cloth".[21] Indeed, we cannot completely ignore which beliefs we hold before becoming fictionalists. If the point of keeping moral thoughts is that they motivate us, the content of the moral fiction must be restricted to what can spark moral motivation, and this will depend on which moral beliefs we are used to. Some instrumentally desirable types of actions are so far removed from traditional conceptions of moral behaviour that we could not have a sense of moral 'must-be-doneness' about them – for example, avoiding tax. Furthermore, it is possible that the specific moral thoughts that correspond to a fictionalist's previous moral beliefs will have the strongest motivational power over them.

However, this does not entail that the content of our moral fiction will be identical to the content of our prior moral beliefs; on the contrary, it is very improbable that we had already accepted precisely the most useful version of morality. It is likely that some of a fictionalist's prior moral beliefs lead to behaviour that is not ultimately advantageous to them, and it is even more likely that the set of prior beliefs does not lead to behaviour that is *maximally* advantageous to them. In choosing the fiction, we need to consider which moral thoughts can give us a sense of 'must-be-doneness', but it is also crucial whether the actions these thoughts would motivate us to do actually maximise our preference-satisfaction. Moral beliefs that are not to the believer's advantage should not be preserved in the fiction. Surely, if we give up on moral truth

[20] Joyce, *The Myth of Morality*, 219.
[21] Nolan, Restall, and West, 'Moral Fictionalism versus the Rest', 327.

and engage with morality purely for the instrumental value of doing so, improvements can be made.

It may well be true that adopting a conservative fiction is more useful than not adopting a moral fiction at all; yet, for Joyce's purposes, this is not enough. His case for fictionalism is not that it is sufficiently advantageous, but that it is the most advantageous policy available to the error theorist. Consequently, he cannot settle for a fiction that is merely good enough. In Joyce's framework, maximising preference-satisfaction is all that matters. A fiction should therefore be chosen if and only if it is optimally useful in this sense; there is no room for other considerations.[22] A recommendation to adopt a fiction that corresponds to our prior moral beliefs rests on a failure to include other possible moral fictions in the calculation of which policy is best.

Joyce does not explicitly defend adopting a fiction that preserves the content of our moral beliefs; his recommendation is rather that error theorists accept the "conceptual framework" of morality.[23] In his words, this merely involves accepting general claims such as "There are obligations and prohibitions" and "Wrong-doers deserve punishment", as well as minimal constrains on the content of moral norms – for example, "Torturing babies to pass the time is always wrong".[24] However, as I have argued, fictionalism comes with a criterion for settling the specific content of the fiction. Recommending a moral fiction in the form of an open-ended conceptual framework is not in line with this. More implicitly, Joyce does suggest that the fiction would have conservative content. Only if fictionalism preserves the traditional content of moral discourse can it play the role, which Joyce ascribes to it, of undermining worries about the practical impact of accepting error theory. Furthermore, Joyce typically describes fictionalism as the recommendation to *continue* with moral discourse, but as a fiction.[25] This implies that adopting fictionalism is merely a matter of transforming an attitude of belief in moral propositions to one of pretence. As an illustration of the

[22] A proponent of fictionalism may want to reject the commitment to maximisation and defend a satisficing conception of rationality instead, which might allow the choice of a suboptimal fiction to be rational. However, it is doubtful that a true satisficing conception of rationality is compatible with a strictly instrumental and preference-based view of reasons (Michael Byron, 'Satisficing and Optimality', *Ethics* 109, no. 1 (1998): 67–93).
[23] Joyce, *The Myth of Morality*, 195.
[24] Joyce, 195.
[25] E.g. Joyce, 221.

transition to fictionalism, Joyce describes a hypothetical person who is raised to think of actions as right or wrong but eventually becomes an error theorist, at which point "these patterns of thought are now so deeply embedded that in everyday life she carries on employing them, and is happy to do so – she becomes a moral fictionalist".[26] A natural interpretation of this is that the content of her moral thoughts is not affected by her transforming from a moral believer into an error theorist and fictionalist.

I believe this is mistaken: what is missing from this story is that once the error theorist decides to carry on with moral thinking for the sake of their preference-satisfaction, they should wonder if they can adjust their patterns of thought to be more advantageous. Unless their moral views were already precisely the most beneficial ones for them, they should make changes. This shows that there are two steps involved in becoming a revolutionary moral fictionalist. What is normally emphasised is the revolution in one's *attitude* towards morality: adopting fictionalism means switching from believing moral propositions to pretending that they are true. However, it follows from the argument for fictionalism that a second revolution is involved: a revolution in the *content* of morality. When the fictionalist switches from belief to pretence, they should also change the content of the moral propositions they accept in this way.

2. Fictionalism and contractarianism

Fictionalism is strongest as a policy that is accepted collectively rather than independently. Since the instrumental value of acting in accordance with moral norms largely depends on others reciprocating, the habit of moral thinking is advantageous to an individual under the condition that this practice is widespread in their society. Joyce presents his fictionalism as a way forward for a group of error theorists: "By asking what *we* ought to do I am asking how a *group* of persons, who share a variety of broad interests, projects, ends – and who have come to the realization that morality is a bankrupt theory – might best carry on."[27]

[26] Joyce, 224.
[27] Joyce, 177.

Arguably, such a group will be best advised to adopt a single moral fiction together: if fictionalists do not coordinate their moral thoughts and resulting behaviour, the cooperation their engagement with morality leads to will be suboptimal. For example, pretending that it is wrong to break promises is advantageous for an agent if done collectively, but puts them at risk of exploitation if some members of the group do not accept this prohibition. To see which moral fiction a fictionalist should adopt, then, we need to investigate which moral fiction is best for the members of their society to adopt together. However, there is no room for considerations about the collective good in Joyce's framework; individual preference-satisfaction is all that matters. Therefore, the content of the moral fiction of a society of fictionalists will be restricted by what is advantageous to its members as individuals.

With this in mind, the moral fiction can be understood as a social contract that fictionalists endorse for their own benefit. Interestingly, this means that there is an important resemblance between fictionalism and accounts of morality as a rational agreement between individuals. In particular, Joyce's fictionalism is closely related to David Gauthier's contractarian theory of morality, which aims to vindicate morality as a set of constraints which it is rational for actual persons to agree to and comply with.[28] Note that Joyce and Gauthier are nevertheless opponents: Gauthier is not an error theorist.[29] He assumes that genuine moral requirements exist and provides a theory of why we should comply with them, rather than a theory of what to do in absence of moral facts. Still, we can learn more about fictionalism by looking at contractarianism.[30] As I will show, Gauthier and Joyce justify moral practices along the same lines. For contractarianism, it has long been established what implications this justification has for the content of the vindicated morality. Due to the parallels between Gauthier's and Joyce's arguments, these implications are relevant for fictionalism too.

There are three main elements that make Joyce's fictionalism and Gauthier's contractarianism significantly similar. The first is that both use strictly non-moral premises to argue in favour of adopting moral practices. It is clear that Joyce, as an error

[28] Gauthier, *Morals by Agreement*.

[29] Eline Gerritsen, 'Morality as a Fiction or Agreement: Joyce versus Gauthier' (MLitt dissertation, University of St Andrews & University of Stirling, 2017).

[30] The term 'contractarianism' can be used for both Hobbesian social contract theories and Kantian contractualist theories. In this paper, it refers specifically to Gauthier's Hobbesian theory.

theorist, must argue for embracing moral thinking without referring to moral facts. Gauthier likewise justifies morality on terms that are acceptable to someone who does not already recognise moral demands: "We are committed to showing why an individual, reasoning from non-moral premisses, would accept the constraints of morality on his choices".[31] This agent Gauthier has in mind does not yet distinguish "between what he may and may not do" or "recognize a moral dimension to choice".[32] Therefore, just like Joyce's, Gauthier's justification of morality is characterised by a complete absence of moral considerations.

Related to this is the second important shared element, the strictly instrumental view of reasons. Gauthier dismisses attempts to vindicate morality with a 'moralised' conception of rationality, such as notions of rationality that presuppose an impartial viewpoint.[33] Instead, in Gauthier's view, a person acts rationally if and only if she "seeks the greatest satisfaction of her own interests".[34] Here, a person's interests are understood simply in terms of her preferences. Thus, like Joyce, Gauthier believes that an agent has a reason to perform an action if and only if it maximises their own preference-satisfaction. This gives them the same starting point for a justification of moral practices: for both Joyce and Gauthier, the task is to show that engaging with morality maximises an individual's preference-satisfaction.[35] Any other consideration in favour of morality will not be of the right kind.

Finally, Joyce and Gauthier make a similar case for why morality is justified in this way: central to both accounts is that being moral entails being cooperative, and that cooperation is beneficial to the individual. Where their arguments for these two claims diverge, this is mainly because Gauthier's account is much more elaborate in this respect than Joyce's. Since Joyce mostly leaves open why it is advantageous to be cooperative in interaction with others, Gauthier's account of this can be plugged into Joyce's argument to make the latter more complete. In my view, then, a more sophisticated version of Joyce's justification of moral practices would be even more similar

[31] Gauthier, *Morals by Agreement*, 5.
[32] Gauthier, 9.
[33] Gauthier, 4–8.
[34] Gauthier, 7.
[35] For Joyce, this can include the satisfaction of other-regarding preferences. I address how this affects the content of the fiction in section 3.2.

to Gauthier's. In any case, the contractarian and fictionalist justifications of morality must run largely parallel, due to their shared starting point. With this in mind, I will now discuss Gauthier's justification of morality and its implications for morality's content.

Gauthier claims that morality promotes cooperation by demanding impartial constraints on an agent's utility maximisation. In his view, adopting morality consists of letting go of a policy of directly pursuing what is best for you without restrictions. He argues that restricting yourself is ultimately in your best interest. In the first place, it is beneficial for you if *others* adopt impartial constraints on their behaviour, since they may otherwise pursue their own preference-satisfaction at great cost to you. However, others cannot be expected to constrain their behaviour towards you unless you accept these constraints as well. To access the benefits of cooperation, you must play by the rules. It is therefore in one's best interest to forego a strategy of directly pursuing what is in one's best interest, on the condition that others do the same.[36] Gauthier claims that others will notice if you try to fool them into cooperating with you by merely superficially accepting moral constraints. Therefore, to benefit from moral practices, you must fully endorse them by developing a deeply ingrained disposition to constrain your direct utility-maximisation, including in specific situations in which this is disadvantageous to you.[37]

---

[36] David Gauthier, 'Why Contractarianism?', in *Contractarianism and Rational Choice: Essays on David Gauthier's Morals by Agreement*, ed. Peter Vallentyne (New York: Cambridge University Press, 1991), 23.

[37] Gauthier, *Morals by Agreement*, 172–77. I will take for granted that this argument succeeds. In reality, there are several major problems that Gauthier is known to face and seems unable to overcome. For one, there are many conceivable policies to adopt towards utility-maximisation, and Gauthier has not established that fully committing yourself to constrained utility-maximisation is the best option of all (David Copp, 'Contractarianism and Moral Skepticism', in *Contractarianism and Rational Choice: Essays on David Gauthier's Morals by Agreement*, ed. Peter Vallentyne (New York: Cambridge University Press, 1991), 220–21; Holly Smith, 'Deriving Morality from Rationality', in *Contractarianism and Rational Choice: Essays on David Gauthier's Morals by Agreement*, ed. Peter Vallentyne (New York: Cambridge University Press, 1991), 238–43; Geoffrey Sayre-McCord, 'Deception and Reasons to Be Moral', in *Contractarianism and Rational Choice: Essays on David Gauthier's Morals by Agreement*, ed. Peter Vallentyne (New York: Cambridge University Press, 1991), 181–95. Moreover, even if it is true that it is rational to adopt a disposition of constraining your behaviour, this fails to show that it is rational to *act on* such a disposition (Derek Parfit, *On What Matters: Volume One* (Oxford: Oxford University Press, 2011), 433–47; Smith, 'Deriving Morality from Rationality', 244–49; Copp, 'Contractarianism and Moral Skepticism',

The nature of this justification has important consequences for the scope of moral obligations it supports. In contractarian thought, you have a reason to behave in a considerate way towards others not because they have intrinsic value, but because they are instrumentally valuable to you.[38] However, crucially, not every person provides instrumental value for us in the sense Gauthier is interested in. Not everyone is in a position to engage in mutually beneficial cooperation with you; some relations are unavoidably asymmetrical, and here restrained cooperation can bring one party more costs than benefits. Yet, it follows from Gauthier's argument that if others cannot offer beneficial cooperation in return for you constraining your behaviour towards them, then you do not have a reason for adopting such constraints. He acknowledges that this leads contractarianism away from common conceptions of the content of morality: "we may agree that the moral constraints arising from what are, in the fullest sense, conditions of mutual advantage, do not correspond in every respect to the 'plain duties' of conventional morality".[39]

In particular, Gauthier cannot recognise moral obligations towards several categories of persons that we do normally take ourselves to owe moral consideration to. The first is persons outside of the society that our social interactions take place in. Interactions with outsiders are rare – when you do have an opportunity to benefit them, this is unlikely to result in a benefit to you. Gauthier thus argues that the needs of a different society that one does not cooperate with are morally irrelevant.[40] Non-human animals are a major subcategory of outsiders that do not warrant moral consideration on the contractarian picture.[41]

A second category excluded from the contractarian morality is future persons whose lives do not overlap with yours. The asymmetric relation here is obvious: any

207). An interesting upshot of the similarities with contractarianism is that fictionalism faces these same challenges. Perhaps Joyce is able to provide an answer to Gauthier's problems – for example, he may argue that committing to constrained utility-maximisation is the best available option due to our weakness of will. However, it is important that we are as critical of these steps in the fictionalist argument as we are in the case of the contractarian argument.

[38] Jean Hampton, 'Two Faces of Contractarian Thought', in *Contractarianism and Rational Choice: Essays on David Gauthier's Morals by Agreement*, ed. Peter Vallentyne (New York: Cambridge University Press, 1991), 48.

[39] Gauthier, *Morals by Agreement*, 268.

[40] Gauthier, 282–88.

[41] Gauthier, 268.

constraints you put on your behaviour for the sake of future persons cannot be reciprocated. If no cooperation is possible with our descendants, no injustice is committed when you do not consider their interests, even if this leaves them with an uninhabitable world.[42]

Another striking omission from the contractarian morality are moral obligations towards severely disabled or chronically ill persons: "Only beings whose physical and mental capacities are either roughly equal or mutually complementary can expect to find cooperation beneficial to all. (...) Among unequals, one party may benefit most by coercing the other, and on our theory would have no reason to refrain".[43] Some disabled persons permanently cannot engage in mutually beneficial cooperation, either because of the nature of their disability itself or because public space is not made accessible to them.[44] These persons therefore have no grounds for demanding any kind of moral respect in the contractarian picture. Clearly, Gauthier's instrumental justification of moral practices leaves him with a very limited form of morality.

Because fictionalism has the same starting point as contractarianism, Joyce unavoidably faces very similar implications for the content of morality. Revolutionary fictionalists should accept the specific moral fiction that is maximally useful for their individual preference-satisfaction. Just like Gauthier, Joyce makes the usefulness of moral thoughts and behaviour contingent on the usefulness of cooperation. However, as we learned from contractarianism, a morality purely based on the rewards of cooperation cannot match the extension of ordinary conceptions of morality. Many persons we normally ascribe moral status to do not qualify for mutually beneficial cooperation. If the moral fiction is to be chosen on this basis, then a fictionalist should not pretend that it is morally wrong to harm outsiders, non-human animals, future persons or severely disabled persons.

My claim that fictionalism fails to account for some conventional moral obligations is not meant to imply that there is a single conventional morality that all moral

---

[42] Gauthier, 298. Gauthier argues that there are cooperative links between all generations due to generational overlap (Gauthier, 299). This argument is shown to rest on highly implausible assumptions in Gustaf Arrhenius, 'Mutual Advantage Contractarianism and Future Generations', *Theoria* 65, no. 1 (1999): 25–35.

[43] Gauthier, *Morals by Agreement*, 17.

[44] Martha C. Nussbaum, *Frontiers of Justice: Disability, Nationality, Species Membership* (Cambridge, Mass.: Belknap Press, 2007), 117–18.

believers endorse. Moral beliefs are not uniform. Nonetheless, there appears to be agreement among most people about some aspects of morality – and fictionalism turns out to fall short here. It seems highly unusual to recognise no moral obligations whatsoever towards other societies, non-human animals and future generations. Moreover, the requirement to take care of the most vulnerable can be seen as one of the basic elements of our moral code. The revolutionary fiction goes directly against this if it undermines the moral status of severely disabled members of society.

3. Objections and replies

3.1 The feasibility of changing moral thinking

A potential objection to my account of fictionalism is that we are psychologically incapable of fully adopting the moral fiction with revolutionary content. The worry is that our moral judgements, emotions and dispositions are so deeply ingrained in us that swapping them for a more useful set is not an option. Alternatively, one could object that radically changing our moral thinking would not be most advantageous, even if it is possible. It may be thought that a newly adopted set of moral thoughts would lack the motivational power that is supposed to make them useful. In addition, accepting a revisionary fiction would come with psychological resistance, while a conservative fiction is more user-friendly in comparison. Due to the mental adjustment required, the thought goes, it would be either impossible or disadvantageous to adopt the revised moral fiction I have sketched. How much change is actually possible in the moral lives of a hypothetical group of committed error theorists is an empirical question which we cannot settle here. Still, there are some reasons to expect that a revolution is both possible and most conducive to fictionalists' preference-satisfaction.

Firstly, my argument does not presuppose that fictionalists are capable of endorsing and being motivated by fictional moral obligations they did not previously believe in; my suggestion is merely that a range of traditional obligations would be *left out* of the fiction, without new obligations or rights replacing them. That fictionalists should not pretend that they have obligations towards future persons does not mean that they should pretend that it is morally permissible to ruin future lives; instead, they should

not think about future persons in moral terms at all. Their moral thinking should stay 'turned off' when they are considering actions that affect those persons, just like they do not engage in moral thinking when considering whether to go for a run. This change will still take some mental effort. However, it is an effort that will be worth it: since the obligations that are removed from the fiction are precisely those that constitute an uncompensated burden to many, the fictionalist has a strong motive to do what they can to remove any traces of moral beliefs about them. It seems plausible that many persons would happily stop seeing themselves as obligated to donate to an overseas disaster relief fund, for example. When the fictionalist does revert to traditional moral thoughts, they should try to correct themselves and refrain from acting on them, with an eye to the substantial future benefits of being unburdened by these obligations.

Secondly, fictionalism is a long-term policy. Presumably, the ideal is that groups of error theorists who adopt fictionalism will go on to be immersed in the moral fiction for good. While they may struggle to fully internalise the revised moral content at first, they will have the rest of their lives to get used to the more advantageous moral practices. When the erosion of moral thinking is gradual, the psychological resistance to it will be less strong. New generations, moreover, can be brought up to be committed to the adjusted moral values from the start. Therefore, if the revisionary nature of the fiction I sketched makes it difficult to internalise effectively, or if the mental effort involved in internalising it counterbalances some of its advantage over a conservative fiction, this will plausibly be minimised over time. Nonetheless, this does not make the moral practices that the society of fictionalists would end up with – and which fictionalists should aspire to move towards now – any less radical from our own perspective.

3.2 Other-regarding preferences

Another possible worry about my account of the revolutionary fiction is that it seems to ignore our other-regarding preferences, which might justify more extensive moral practices. Gauthier dismisses social practices that are only beneficial for agents due to their sympathy for others as a form of exploitation.[45] Consequently, his goal is to show

[45] Gauthier, *Morals by Agreement*, 11.

that endorsing moral obligations gives a net benefit to an agent even if we do not take their other-regarding preferences into account. Joyce need not limit himself in this way: he can defend fictionalism as the best way forward for error theorists given their full range of preferences, not just their selfish ones.

However, it is unlikely that fictionalists' other-regarding preferences can support all traditionally recognised moral obligations; the average agent's desires to help others are too limited. Empirical research has shown that purely altruistic behaviour is normally motivated by feelings of empathy.[46] Here, empathy is understood as an "another-oriented emotional response elicited by and congruent with the perception of another person in need".[47] When we do accommodate our capacity for empathy, we must also acknowledge its limits. Firstly, empathy is subject to a 'familiarity bias': we are more likely to empathise with persons who are similar or in some way personally connected to us. This bias applies to friends and family as well as to members of one's own ethnic or racial group.[48] Secondly, there is evidence of a 'here-and-now' bias: we are more likely to have an emphatic response towards a person who is physically present or salient in some way that draws our immediate attention.[49] To illustrate, seeing the photo of a specific drowned refugee all over the news will evoke feelings of empathy and an accompanying willingness to help, which is absent when large numbers of similar victims are only presented as anonymous statistics. Relatedly, Jesse Prinz points out that "empathy is hard to evoke for foreign masses".[50] Empathy focuses on immediate, local needs, and ignores more widespread or systematic problems.[51] Strikingly, these limits to our capacity for empathy largely map onto the gaps in the

[46] C. Daniel Batson, Nadia Ahmad, and E. L. Stocks, 'Four Forms of Prosocial Motivation: Egoism, Altruism, Collectivism, and Principlism', in *Social Motivation*, ed. D. Dunning (New York: Psychology Press, 2011), 111. See also C. Daniel Batson et al., 'Is Empathic Emotion a Source of Altruistic Motivation?', *Journal of Personality and Social Psychology* 40, no. 2 (1981): 290–302; Dennis Krebs, 'Empathy and Altruism', *Journal of Personality and Social Psychology* 32, no. 6 (1975): 1134–46.

[47] Batson, Ahmad, and Stocks, 'Four Forms of Prosocial Motivation', 110.

[48] Martin L Hoffman, *Empathy and Moral Development: Implications for Caring and Justice* (Cambridge: Cambridge University Press, 2000), 206–9. See also Jesse Prinz, 'Against Empathy', *The Southern Journal of Philosophy* 49 (2011): 227–28.

[49] Hoffman, *Empathy and Moral Development*, 209–13.

[50] Prinz, 'Against Empathy', 224.

[51] Prinz, 228.

contractarian morality. As a result of both the familiarity bias and the here-and-now bias, we normally do not feel a substantial level of empathy towards future persons or strangers with no relation to us, including non-human animals. Consequently, we are unlikely to have a preference to help them for their own sake.

Furthermore, while selfish behaviour is normally kept in check by moral beliefs, fictionalists will lack such a constraint. A moral believer who does not feel empathy for a group of strangers may nonetheless prefer to help them because they believe that is the right thing to do. This type of motivation is not available to a fictionalist, and is therefore irrelevant for determining the content of the moral fiction. The same is true for our preferences to help others that stem from a preference to comply with moralised social norms: a society of fictionalists should get rid of such norms when they are at odds with the most useful version of morality. Again, even if these norms remain internalised by fictionalists at first, the long duration of the fictionalist project ensures that they can eventually be phased out. Then, there will be no social punishment to fear when you fail to help people who cannot reciprocate. Hence, some of the motives people currently have to take the needs of others into consideration do not apply to fictionalists. We can conclude that the most useful moral fiction will not include all traditional moral obligations even if we take fictionalists' full range of preferences into account.

With respect to moral obligations towards severely disabled persons, taking empathy-based preferences into account may have a better effect. The average abled person may be close enough to at least some disabled persons for her empathy to be triggered by their needs. If so, moral obligations towards disabled persons can be justified with reference to fictionalists' other-regarding preferences. Still, the foundation of these moral obligations would be very fragile: if it turns out that not enough fictionalists care about persons who cannot offer beneficial cooperation, there is no ground for extending moral consideration to them.

Of course, there is no doubt that the preferences of fictionalists who are disabled themselves will give them good reason to endorse obligations towards disabled persons. These preferences are not to be ignored. However, as I have argued, implementing a moral fiction should be treated as a collective enterprise, and its content will depend on what other fictionalists have reason to accept. The needs of vulnerable persons carry no extra weight here and, when they are a minority, can easily fail to be

reflected in the moral obligations that fictionalists are collectively willing to accept. Preference-based cost-benefit analyses of the kind Gauthier and Joyce use just cannot produce a conventional morality.

3.3 So what?

Faced with my sketch of the revolutionary content of the moral fiction, the advocates of fictionalism may respond: 'So what?'. In the case of Gauthier, the result that his theory breaks with conventional moral views directly suggests that it is false: unless our moral beliefs are strongly mistaken, the contractarian picture of morality is inaccurate.[52] In contrast, because revolutionary fictionalism is not a theory of the moral facts, its revisionary implications do not show it to be false. Those who are already fully committed error theorists may therefore see the discrepancies between the moral fiction and conventional moral beliefs as an interesting yet unproblematic result. Indeed, a Nietzschean nihilist may even embrace the 'death of morality' as an opportunity for us to realise that "many other, above all higher, moralities are possible or ought to be possible"[53] – now, in the form of a fiction.

I grant that, for all I have said in this chapter, error theory could still be true and fictionalism could be the best way to move forward. Even so, fictionalism's revolution in moral content does constitute a problem for both. Joyce presents fictionalism as a strategy that can invalidate concerns about the disruption that collectively accepting error theory might bring.[54] If it were to succeed in fulfilling this promise, that would significantly improve the status of error theory. Having radical implications is a disadvantage of a theory, even if those who are already on board are willing to accept them. Error theory is an unattractive position if it undermines both our common-sense beliefs and our way of life. Hence, some argue that a good proposal for what to do with morality after embracing error theory must let us keep practices that seem important to us, thereby making error theory more palatable.[55] Achieving this stability is

[52] Walter Sinnott-Armstrong, *Moral Skepticisms* (Oxford: Oxford University Press, 2006), 170–71.
[53] Friedrich Nietzsche, *Beyond Good & Evil: Prelude to a Philosophy of the Future*, trans. Walter Kaufmann (New York: Vintage Books, 1989), 202.
[54] Joyce, *The Myth of Morality*, 231.
[55] Lutz, 'The "Now What" Problem for Error Theory'.

supposed to be one of the main strengths of fictionalism. Therefore, it is highly problematic that it turns out to fail in exactly this respect. The moral fiction that should be chosen, based on its value defined in terms of individual preference-satisfaction, would significantly restrict our moral practices. This amplifies the familiar counter-intuitiveness objection to error theory: combined with fictionalism, error theory not only entails that all moral beliefs are mistaken, but also contradicts our social practices and judgements on how to behave towards others. It is not the case that adopting fictionalism allows a society of error theorists to go on as before. Consequently, it cannot be employed to remove the burden of having revolutionary practical implications from error theory.

While fictionalism is only one possible way forward after giving up on moral facts, error theorists cannot simply avoid the problem I have raised by opting for one of the alternatives. Abolitionism clearly reinforces practical concerns about error theory, because it involves a complete disruption of our moral practices by definition. To alleviate these concerns, the abolitionist would need to provide a very convincing story about why, on the whole, we should not fear but welcome this disruption. That conservationism may also imply a revision of thought and behaviour is perhaps less obvious. However, it is important that, after accepting error theory, conservationists continue with having moral beliefs purely for their instrumental value. Given the assumption that there are no true moral beliefs, the content of these beliefs is not to be determined in the normal way, by the aim of appropriately responding to the evidence and representing the facts. Instead, conservationists should arguably make themselves have the moral beliefs that are most useful to them. If so, the implications I have drawn for fictionalism can be repeated for conservationism. While there may be relevant differences between believing and make-believing, I suspect that the most useful moral beliefs will be very similar in content to the most useful moral fiction. Therefore, my conclusion that fictionalism has revolutionary practical implications uncovers a more general problem for Joyce's form of moral error theory: we cannot consistently go on as before when instrumental value is all we have left. This does not entail that error theory is false, but does make it an even more radical and unattractive position than it would otherwise be.

4. Conclusion

Revolutionary fictionalism is more radical than has been assumed, due to its overlooked implications for the content of morality. Because error theorists are supposed to adopt a fictional morality purely for its instrumental value, they should choose the specific fiction that has most instrumental value for them, rather than the fiction that corresponds to their prior moral beliefs. This leads to the second revolution of fictionalism: although fictionalism is presented as merely requiring a change in attitude, it also requires a change in the set of moral obligations we accept. I have argued that the most useful moral fiction can be expected to exclude obligations to persons outside our society, non-human animals and future persons, and to make the moral status of severely disabled persons contingent on our empathy. This substantially limits the scope of moral thinking in comparison to ordinary moral beliefs, leaving fictionalists with impoverished moral practices. As a result, fictionalism fails to save error theory from the burden of having radical practical consequences. In the end, the comforting idea that we could coherently respond to moral error theory by simply transforming our moral beliefs into make-belief appears to be a mere fiction.

It may be thought that my discussion of fictionalism is in tension with my defence of Humean theories of authority. I emphasised in several chapters that the vast majority of real human beings will have commitments that make morality authoritative for them. Although this may seem to contradict the claim that fictionalists would not have preference-based reasons to endorse all conventional moral obligations, there are important differences between these claims. Firstly, my claim about fictionalists is specific to a society where moral error theory is collectively accepted. Here, as I explained, moral norms cannot be justified by preferences shaped by moral beliefs or social expectations to comply with traditional moral norms. For real persons outside of a society of fictionalists, however, such preferences are in play. Our commitments can include a genuine concern with doing the right thing. In addition, we may have a preference to refrain from harming outsiders, severely disabled persons, and future generations on the basis that we consider them to be persons we owe respect to. Moralised social expectations also have an important role in justifying compliance with morality for non-fictionalists: since the traditional moral norms do have currency in our society, we are likely to face sanctions and frustrate even our selfish

preferences when we do not comply with them. A second difference is that the question of what reasons we have to obey a moral fiction directly determines the correct content of that fiction. Our reasons (or lack thereof) to be moral arguably do not affect the content of the true moral norms in the same way. My discussion of the authority of morality has taken the content of morality as fixed. Given that the moral norms are what they are – which includes obligations to those left out of the revolutionary fiction – I have argued that most of us have authoritative reasons to comply with them. For fictionalists, only a subset of moral norms are justified, because they can discard the ones that do not maximise their individual preference-satisfaction. We cannot pick and choose in this way with a non-fictional morality. The question I have therefore focused on, in previous chapters, is whether morality is justified as a whole – I maintain that, for nearly all of us, it is.

Conclusion

In this thesis, I explored how we can make sense of normativity in a way that does not lead to an error theory about either normativity or morality. We are used to making normative judgements about the reasons we have and the actions we ought to choose. Normative error theory denies that any such judgements are true, making it a highly radical position. The same applies to moral error theory: as I argued in the last chapter, it not only contradicts deeply held beliefs and strong intuitions, but also has – at least in combination with fictionalism – revolutionary implications for our moral practices. With this much at stake, it is important to examine the assumptions about normativity and morality in the background of these theories.

My focus lies in particular on the conceptual commitments involved in normative and moral error theory. I have not argued against the metaphysics of the error theorist; I share a scepticism about the existence of *sui generis*, non-natural and irreducible normative properties. To conclude from a rejection of that type of property that we must give up on normative truth, however, is to throw the baby out with the bathwater. We can deny that there are irreducibly normative properties without denying that there are normative or moral properties – as I have argued, irreducibility is not essential even to the authoritative form of normativity. Non-naturalism and error theory are among the most prominent theories in present-day metaethics. This may seem surprising: given that the standard forms of non-naturalism are committed to the existence of robust, distinctly normative properties, while error theory denies normative properties to exist, how can they both receive so much support? Yet, while non-naturalism and error theory lie on opposite sides of the spectrum when it comes to their

metaphysics, they have their understanding of normativity in common. Both conceive of normative properties as utterly distinct from plain natural properties. At least for the prominent Mackie-inspired strand of error theory I have discussed, it is specifically this conception of them that leads to the rejection of normative properties. In my view, the most promising strategy for avoiding an error theory is not to argue that irreducibly normative properties do exist, but to argue that we do not need them; we can conceive of normativity and morality as having a different, less mysterious nature. We need to be wary of taking the interpretation of normativity of non-naturalists and error theorists for granted and, instead, have a more open debate on what counts as the normativity we are trying to vindicate.

In chapters two through four, I showed that there is room for an alternative to the non-naturalist understanding of authoritative normativity. The first step was to elucidate the distinction between the formal and authoritative forms of normativity. Recognising this distinction is important for getting clear on the normativity we are trying to account for. It is specifically authoritative normativity which is considered, by non-naturalists and error theorists but also some naturalists, to resist a naturalistic reduction. I have rejected this common view and shown the way to an alternative understanding of authority.

In the second chapter, I argued that there is conceptual room for metaphysically naturalist theories that do not treat authoritative normativity as an irreducible property that is intrinsic to norms. It is essential to authoritative normativity that it has a normative significance over and above that of formal normativity; whether this significance can be accounted for by purely natural facts is an open question. In chapter three, I provided a norm-based framework for theories of authoritative normativity. I understand authoritative normativity to be grounded in authoritative norms, where norms are authoritative if and only if they are justified. My overview of different types of views on the justification of norms emphasised that treating normative authority as irreducible is far from the only option. I showed, in particular, how we can explain normative authority as arising out of facts about our identity, practical standpoint, desires, or collective flourishing. In chapter four, I explored the implications and potential problems of the Humean category of theories, which reduce authoritative normativity to natural facts about agents' contingent commitments. My opponent may think that even if Humean theories are not ruled out by the concept of authoritative

normativity, substantial objections to them still show them to be false. Two pressing concerns are that Humean theories make the authority of norms a contingent matter, and that they do not respect the is/ought gap. I argued that these broad objections do not establish that a Humean view of authority does not work; insofar as it does diverge from our intuitive understanding of authority – and not merely from a non-naturalist interpretation of it – this cost can be outweighed by the problems of non-naturalism that the Humean avoids. Therefore, Humean theories offer a valid and promising way to make sense of authoritative normativity.

I have not established that a Humean theory of authoritative normativity is correct, nor have I defended a specific one as the most successful. Instead, I conclude that Humean theories form a legitimate alternative to a non-naturalist understanding of authority, and thereby deserve serious consideration. Of course, Humean theories of normativity are not new, and they have traditionally been the subject of discussion in metaethics. In the more specific debate on authoritative normativity that has recently expanded, however, they lack this attention. It is assumed that reductive theories could only capture formal normativity, not the authoritative form of normativity we are interested in. This obscures the fact that that Humean theories can provide an answer to scepticism about the existence of authoritative normativity. Normative error theorists need to show that we must conceive of authoritative normativity as something mysterious, instead of understanding it to arise out of unproblematic facts about our commitments. This requires more substantial work than it normally receives; Humean options cannot be put to the side by simply identifying authoritative normativity with irreducible normativity. The main substantive arguments against Humean theories are not decisive either. In contrast to what is commonly thought, a form of normative significance that is wholly explained by and contingent on agents' commitments could indeed count as the normative authority we are looking for. To deny that any such account is correct, error theorists and other opponents must engage more elaborately with the details of individual Humean theories.

In the same way normative error theory is based on an unnecessarily demanding understanding of authoritative normativity, moral error theory relies on an implausibly narrow conceptual thesis. We think of moral norms as especially important and forceful. Yet, this does not entail that morality must necessarily have a normative authority that does not depend on agents' commitments. In chapter one, I argued that

'why be moral?' is a meaningful question which asks for authoritative reasons to do what morality requires. It is not self-evident that there are such reasons; in contrast to what Prichard suggests, this issue is up for a fruitful philosophical debate. As I discussed in chapter five, the conceptual thesis of Olson and Joyce denies this. They take everyday moral claims to come with a conceptual commitment to categorical authoritative normativity, such that no moral claims can be true if that type of normativity does not exist. We should not give up on moral truth easily; especially considering the practical disruption of collectively accepting moral error theory that I discussed in chapter six, we must critically examine the understanding of moral properties behind the theory that they do not exist. The conceptual thesis does not survive this examination – it is far from clear that categorical authority is a non-negotiable element of everyday moral concepts. This means that we can offer a commitment-dependent account of authoritative normativity, as I have done, without thereby losing moral properties.

This thesis raises several questions that can form the subject of further research. For one, I have assumed that we can take the content of norms as fixed when we ask whether they are authoritative. My intention is to offer an account specifically of the authority of norms, which leaves open what norms are constituted by and how their content is determined. We can question, however, to what extent these issues can be separated – some theories of the authority and content of norms may be incompatible. Take the Humean view of authority. It would be desirable for the Humean to say that, while the authority of norms depends on agents' commitments, the content of norms need not be mind-dependent – this would avoid relativism about moral requirements, for example. However, this raises the challenge of showing how this combination of views can work. Since its metaphysical naturalism is one of the major advantages of the Humean view of authority, the Humean will want to avoid a non-naturalist explanation of necessary norms. On the other hand, if we explain norms with natural facts about conventions, for example, we face an element of relativism. To settle this issue, we need a comprehensive account of the metaphysics of norms. This should include attention to how systems of norms are demarcated, how they relate to each other, and whether there is a fundamental division between conventional and non-conventional norms. When we have a better grasp of what norms are, we can in turn provide a more detailed account of their authority.

A further issue I have left open in my discussion of authoritative normativity concerns the weight of reasons. I have asked what makes some norms have authority while others lack it, which suggests that the formal/authoritative distinction puts norms into two separate categories. We may think that, in reality, authority comes in degrees. Even if both morality and etiquette matter, unlike an arbitrary set of rules I have just made up, the requirements of morality seem to have more weight than those of etiquette. Even if etiquette reasons are authoritative for us, they will often lose out to moral reasons. This connects to the relation between authoritative normativity and the ought *simpliciter*. I have said that only the requirements of authoritative norms feed into the ought *simpliciter*. However, the requirements of authoritative norms can conflict. It seems that if all authoritative norms are equally binding and we are not required – outside of the perspective of particular norms – to prioritise any over the others, there is no answer as to what we ought to do all things considered. The account we give of authoritative normativity will inform whether, and how, we can explain the difference in weight of authoritative reasons. On a Humean account, we can explain the relative authority of norms as a matter of the degree to which they are supported by agents' commitments. If norms have authority in virtue of being justified by our commitments, norms that receive more support from those commitments than others will plausibly have more authority. By containing a criterion for prioritising types of norms in this way, a Humean view of authoritative normativity can help us to make sense of the ought *simpliciter*. This is a good illustration of how the norm-based account of authoritative normativity I have developed can be supplemented, extended, and filled in with more details, enabling us to demystify further puzzles of normativity.

Nederlandse samenvatting

We leven in de veronderstelling dat we ons op een bepaalde manier behoren te gedragen. We behoren vriendelijk te zijn, mensen in nood te helpen en ons aan onze beloftes te houden. Daarnaast behoren we goed uit te kijken voor we een straat oversteken, mensen op de juiste manier te begroeten en onze gasten niet te beledigen. Er zijn allerlei voorschriften van toepassing op ons gedrag, die vaak invloed hebben op wat we daadwerkelijk doen. Hoewel deze ervaring van normativiteit een alledaags fenomeen is, vormt ze een filosofisch mysterie: waar komt dit 'behoren' vandaan? Handelingen kunnen vereist, verboden of toegestaan zijn. We kunnen redenen hebben om iets te doen. Dit zijn normatieve eigenschappen van handelingen. Hoe kunnen we normatieve eigenschappen verklaren? Hoe passen deze in de wereld?

In filosofisch onderzoek naar normativiteit krijgen morele voorschriften traditioneel de meeste aandacht. We mogen niet zomaar doen wat we willen; we lijken gebonden aan morele normen die bepaald gedrag verbieden of juist eisen. Dit roept fundamentele vragen op over onze redenen om ons aan deze normen te houden. Wat geeft mij bijvoorbeeld een reden om geld te doneren aan oorlogsvluchtelingen? Hoe komt het dat ik verplicht ben een goed mens te zijn? Waarom ben ik niet vrij om schade toe te brengen aan een ander, ook als het zeker is dat ik hier niet voor gestraft zal worden? De puzzel van normativiteit gaat echter verder dan vragen over moraliteit. Naast morele normen zijn er ook veel andere soorten normen die eisen opleggen aan ons gedrag of onze opvattingen, zoals de wet, schaakregels, etiquettenormen en de regels van de logica. Sommige normen bepalen daadwerkelijk wat we behoren te doen en mogen niet zomaar genegeerd worden. Dit normatieve gezag vraagt om een verklaring.

In de metanormatieve literatuur is er een sterke consensus dat normativiteit niet gereduceerd kan worden tot natuurlijke eigenschappen. Met 'natuurlijke eigenschappen' wordt hierbij grofweg het soort eigenschappen bedoeld dat door de empirische wetenschappen bestudeerd wordt. Voorbeelden van natuurlijke eigenschappen van een handeling zijn dat deze tot meer welzijn leidt of bijdraagt aan de doelen van degene die het doet. Als normatieve eigenschappen uiteindelijk natuurlijke eigenschappen zijn, passen ze binnen ons moderne wetenschappelijke wereldbeeld zonder hier iets mysterieus aan toe te voegen. Veel filosofen denken echter dat zo'n reductie niet mogelijk is, waardoor twee alternatieve zienswijzen over normativiteit in beeld komen. Het eerste is non-naturalisme. Volgens deze groep van theorieën zijn normatieve eigenschappen geen natuurlijke eigenschappen, maar bestaan ze desondanks. Ze voegen dus een speciaal normatief domein toe aan de wereld. Dit maakt zowel de metafysische aard van normatieve eigenschappen als onze kennis ervan mysterieus. De tweede optie is de normatieve vergissingstheorie. Volgens deze theorie bestaan normatieve eigenschappen niet. Alle claims over wat we behoren te doen, niet mogen doen of een reden hebben om te doen rusten daarom op een vergissing en zijn niet waar. Dat maakt dit een erg radicale theorie, die normativiteit niet verklaart maar ontkent. Daarnaast voedt het de *morele* vergissingstheorie. Volgens deze theorie bestaan morele eigenschappen niet; niets is moreel goed of fout, toegestaan of verboden. De normatieve en morele vergissingstheorieën zijn compatibel met ons wetenschappelijke wereldbeeld, maar gaan lijnrecht in tegen onze morele intuïties en alledaagse ervaring van normativiteit. Non-naturalisme en de normatieve vergissingstheorie hebben daarom beide grote nadelen.

Dit proefschrift biedt een alternatief voor non-naturalisme en vergissingstheorieën door te laten zien dat normativiteit niet mysterieus hoeft te zijn. Hiermee breng ik normatieve eigenschappen terug naar het natuurlijke domein. Mijn onderzoeksvraag, die expliciet het onderwerp van sommige hoofdstukken is en de achtergrond vormt voor het hele proefschrift, is als volgt: hoe kunnen we normativiteit begrijpen op een manier die niet tot een vergissingstheorie leidt? In tegenstelling tot non-naturalisten is mijn strategie niet om mysterieuze eigenschappen te accepteren, maar om ze te ontrafelen: ik trek de conceptuele aannames achter de vergissingstheorie in twijfel. Ik beargumenteer dat het niet nodig is om normativiteit te zien als iets dat te vreemd is om te

bestaan – er is ruimte voor een interpretatie van normatieve eigenschappen waarbij deze niet metafysisch of epistemologisch problematisch zijn.

Een belangrijke rode draad die door het proefschrift loopt is het onderscheid tussen formele normativiteit en autoritatieve normativiteit. Formele normativiteit is de vorm van normativiteit die alle normen hebben, op grond van hun vermogen om gedrag of opvattingen voor te schrijven. Autoritatieve normativiteit is een meer substantiële en exclusieve eigenschap: normen met deze vorm van normativiteit hebben normatief 'gezag' en behoren daadwerkelijk gevolgd te worden. Het eerste hoofdstuk introduceert dit onderscheid vanuit de context van de klassieke vraag 'heb ik een reden om te doen wat moreel juist is?' Prichard beweert dat het geen zin heeft om deze vraag te stellen, omdat het antwoord vanzelfsprekend is. Scanlon's interpretatie van het probleem is dat we óf een niet-morele reden geven om het juiste te doen, wat een verkeerd soort reden is, óf aangeven dat er een morele reden is om het juiste te doen, wat een tautologie is. We kunnen dit dilemma echter ontwijken door een onderscheid te maken tussen formele redenen en autoritatieve redenen. Terwijl alle normen verbonden zijn aan formele redenen is het niet vanzelfsprekend dat we een autoritatieve reden hebben om een bepaald soort norm op te volgen. Omdat 'heb ik een reden om te doen wat moreel juist is?' naar autoritatieve redenen vraagt, concludeer ik dat het toch zinvol is om onderzoek te doen naar deze vraag, net als naar normativiteit in het algemeen.

Hoofdstuk twee biedt een conceptuele analyse van autoritatieve normativiteit. Ik laat zien hoe het contrast tussen formele en autoritatieve normativiteit is beschreven in de literatuur en bekritiseer de verschillende diagnoses die zijn gegeven van wat autoritatieve normativiteit anders maakt. Daarnaast richt ik mijn aandacht op het verschil tussen de concepten van autoritatieve normativiteit en 'behoren *simpliciter*'. In de rest van het hoofdstuk ga ik in tegen het prominente idee dat autoritatieve normativiteit een intrinsieke eigenschap van normen moet zijn die niet te reduceren is tot een niet-normatieve eigenschap. Dit idee staat aan de basis van Olson's normatieve vergissingstheorie: het interpreteert normatief gezag als iets dat moeilijk verklaarbaar is. We hoeven deze interpretatie echter niet te accepteren; er is conceptuele ruimte om autoritatieve normativiteit te verklaren met natuurlijke eigenschappen.

Hoofdstuk drie gaat verder in op welke alternatieve interpretaties mogelijk zijn. Eerst formuleer ik de nog abstracte theorie dat autoritatieve normativiteit voortkomt uit gerechtvaardigde normen. Op basis hiervan zijn meer specifieke theorieën te

onderscheiden, die verschillen in hun beeld van de relevante soort rechtvaardiging. Het overzicht van deze theorieën maakt duidelijk dat het veronderstellen van een mysterieuze intrinsieke rechtvaardiging zeker niet de enige optie is. Ik richt me in het bijzonder op een scala aan metafysisch naturalistische theorieën, waaronder Kantiaans constructivisme en meerdere Humeaanse opties. Deze theorieën laten zien hoe autoritatieve normativiteit verklaard zou kunnen worden door middel van natuurlijke feiten over onze identiteit, subjectieve waarden, verlangens of collectief welzijn, wat ons een belangrijk alternatief geeft voor de non-naturalistische interpretatie achter de vergissingstheorie.

In hoofdstuk vier ligt de focus op Humeaanse theorieën die autoritatieve normativiteit verklaren door middel van onze contingente verlangens of waarden. Ik bespreek wat deze theorieën impliceren over welke normen autoritatieve normativiteit (kunnen) hebben, in het bijzonder de implicatie dat conventionele normen niet slechts formeel normatief hoeven te zijn. Daarnaast verdedig ik Humeaanse theorieën tegen de belangrijkste bezwaren. Ik beweer dat de contingentie die aan deze theorieën verbonden is geminimaliseerd kan worden en geen onacceptabel nadeel is, zeker als we de problemen van concurrerende theorieën in gedachten houden. Een andere zorg is dat Humeaanse theorieën een schijnbaar onoverkomelijke kloof tussen het normatieve en het descriptieve proberen te overbruggen. Het bestaan van deze kloof kan echter in twijfel getrokken worden. Ten slotte laat ik zien hoe Humeanen kunnen verklaren dat we behoren te doen wat nodig is om onze doelen te bereiken, wat cruciaal is voor hun interpretatie van het gezag van normen.

In hoofdstuk vijf richt ik me op de morele vergissingstheorie. Argumenten voor deze theorie bestaan uit een conceptuele premisse, die een bepaald onontkoombaar element aan morele concepten toeschrijft, en een substantieve premisse, die ontkent dat dit element echt bestaat. Mijn strategie is om de vergissingstheorie te weerleggen door de conceptuele premisse tegen te spreken. Ik bespreek de conceptuele premisse van Mackie, Olson en Joyce, waarbij helder wordt dat Olson en Joyce een categorische vorm van autoritatieve normativiteit aan moraliteit toeschrijven. Hoewel het plausibel is dat morele normen categorisch zijn in hun voorschriften, is er onvoldoende bewijs voor de specifieke claim van Olson en Joyce, die verder gaat dan dit. Het is niet duidelijk dat het een onontkoombaar element van onze alledaagse morele concepten is dat we onafhankelijk van onze verlangens en doelen altijd autoritatieve redenen hebben

om het juiste te doen. Dit betekent dat het niet klopt dat er geen morele eigenschappen kunnen zijn als er geen categorische autoritatieve normativiteit is.

Het laatste hoofdstuk behandelt de maatschappelijke implicaties van de morele vergissingstheorie: wat zouden we met onze morele oordelen en morele praktijk moeten doen als al onze morele claims inderdaad op een vergissing blijken te berusten? Volgens Joyce's revolutionaire fictionalisme moeten we doen alsof morele eigenschappen toch bestaan, om zo de positieve effecten van moreel denken te behouden. Hij beweert dat dit voorkomt dat de vergissingstheorie radicale praktische implicaties heeft. Ik laat echter zien dat het toepassen van fictionalisme onze morele praktijk niet zou beschermen, maar juist zou ontwrichten. Fictionalisme bouwt op een vergelijkbare rechtvaardiging van morele normen als Gauthier's contractarianisme en heeft daardoor soortgelijke grote gevolgen voor de inhoud van die normen. Omdat fictionalisten alleen in morele termen denken vanwege het persoonlijk nut hiervan, behoren ze in hun morele fictie uitsluitend plichten te accepteren die in hun eigen voordeel zijn. Deze beperking maakt fictionalisme, en daarmee de morele vergissingstheorie, radicaler dan gewenst.

Curriculum Vitae

Eline Gerritsen was born in Winterswijk, the Netherlands on 31 March 1995. After completing her *gymnasium* education in 2013, Eline started a bachelor's degree in Philosophy at Leiden University. She combined this with the Honours College programme 'Expertise in Practice: Advising on Dilemmas in Society'. As part of this, she acted as a research assistant in the university's Centre for Public Values and Ethics. Having visited the University of St Andrews for one semester, Eline graduated *cum laude* from her BA in 2016. She then returned to Scotland for an MLitt in Moral, Political and Legal Philosophy of the St Andrews and Stirling Graduate Programme in Philosophy. Here, she developed her interest in the field of metaethics and started to focus in particular on moral normativity and error theory, resulting in her MLitt dissertation 'Morality as a Fiction or Agreement: Joyce versus Gauthier'. She moved on to doctoral research on normativity more broadly in 2017. A research visit to the Netherlands led her to eventually land in a joint PhD programme with the University of Groningen in addition to the universities of St Andrews and Stirling. During her PhD, she undertook a research internship at the Carnegie Trust for the Universities of Scotland. Additionally, she taught tutorials and lectures on a range of topics in philosophy at the universities of St Andrews, Stirling, and Bayreuth. After her PhD defence, Eline will start a postdoctoral fellowship at Universität Hamburg. Her paper 'The Second Revolution of Moral Fictionalism' is forthcoming in *Ergo*.

Bibliography

Arrhenius, Gustaf. 'Mutual Advantage Contractarianism and Future Generations'. *Theoria* 65, no. 1 (1999): 25–35. https://doi.org/10.1111/j.1755-2567.1999.tb00112.x.

Baker, Derek. 'Skepticism About Ought Simpliciter'. In *Oxford Studies in Metaethics 13*, edited by Russ Shafer-Landau, 230–51. Oxford: Oxford University Press, 2018. https://doi.org/10.1093/oso/9780198823841.003.0011.

———. 'The Varieties of Normativity'. In *The Routledge Handbook of Metaethics*, edited by Tristram McPherson and David Plunkett, 567–81. New York: Routledge, 2017.

Batson, C. Daniel, Nadia Ahmad, and E. L. Stocks. 'Four Forms of Prosocial Motivation: Egoism, Altruism, Collectivism, and Principlism'. In *Social Motivation*, edited by D. Dunning, 103–26. New York: Psychology Press, 2011.

Batson, C. Daniel, Bruce D. Duncan, Paula M. Ackerman, Terese Buckley, and Kimberly Birch. 'Is Empathic Emotion a Source of Altruistic Motivation?' *Journal of Personality and Social Psychology* 40, no. 2 (1981): 290–302. https://doi.org/10.1037/0022-3514.40.2.290.

Berman, Mitchell N. 'Of Law and Other Artificial Normative Systems'. In *Dimensions of Normativity: New Essays on Metaethics and Jurisprudence*, edited by David Plunkett, Scott Shapiro, and Kevin Toh, 137–64. Oxford: Oxford University Press, 2019.

Bicchieri, Cristina. *The Grammar of Society: The Nature and Dynamics of Social Norms*. New York: Cambridge University Press, 2006.

Brink, David O. 'A Puzzle About the Rational Authority of Morality'. *Philosophical Perspectives* 6 (1992): 1–26. https://doi.org/10.2307/2214236.

———. *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press, 1989.

———. 'Moral Realism and the Sceptical Arguments from Disagreement and Queerness'. *Australasian Journal of Philosophy* 62, no. 2 (1984): 111–25. https://doi.org/10.1080/00048408412341311.

Broome, John. 'Is Rationality Normative?' *Disputatio* 2, no. 23 (2007): 161–78.

———. *Rationality through Reasoning*. Oxford: Wiley-Blackwell, 2013.

———. 'Requirements'. In *Homage à Wlodek: Philosophical Papers Dedicated to Wlodek Rabinowicz*, edited by Toni Rønnow-Rasmussen, Björn Petersson, Jonas Josefsson, and Dan Egonsson, 2007. https://www.fil.lu.se/hommageawlodek.

Byron, Michael. 'Satisficing and Optimality'. *Ethics* 109, no. 1 (1998): 67–93. https://doi.org/10.1086/233874.

Calhoun, Cheshire. 'The Virtue of Civility'. *Philosophy & Public Affairs* 29, no. 3 (2000): 251–75.

Case, Spencer. 'Normative Pluralism Worthy of the Name Is False'. *Journal of Ethics and Social Philosophy* 11, no. 1 (2016): 1–20.

Chappell, Sophie-Grace, and Nicholas Smyth. 'Bernard Williams'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Stanford University, fall 2018. https://plato.stanford.edu/archives/fall2018/entries/williams-bernard/.

Christensen, Kasper Højbjerg. 'On The Validity of a Simple Argument for Moral Error Theory'. *International Journal of Philosophical Studies* 24, no. 4 (2016): 508–17. https://doi.org/10.1080/09672559.2016.1203979.

Copp, David. 'Contractarianism and Moral Skepticism'. In *Contractarianism and Rational Choice: Essays on David Gauthier's Morals by Agreement*, edited by Peter Vallentyne, 196–228. New York: Cambridge University Press, 1991.

———. 'Moral Naturalism and Three Grades of Normativity'. In *Morality in a Natural World: Selected Essays in Metaethics*, 249–83. Cambridge: Cambridge University Press, 2007.

———. *Morality, Normativity, and Society*. Oxford: Oxford University Press, 2001.

Côté-Bouchard, Charles, and Clayton Littlejohn. 'Knowledge, Reasons, and Errors About Error Theory'. In *Metaepistemology*, edited by Christos Kyriacou and Robin McKenna, 147–71. Cham: Springer International Publishing, 2018. https://doi.org/10.1007/978-3-319-93369-6_7.

Cuneo, Terence. 'Destabilizing the Error Theory'. In *Epistemic Reasons, Norms and Goals*, edited by Martin Grajner and Pedro Schmechtig, 71–94. Boston: De Gruyter, 2016. https://doi.org/10.1515/9783110496765-005.

Cuneo, Terence, and Sean Christy. 'The Myth of Moral Fictionalism'. In *New Waves in Metaethics*, edited by Michael Brady, 85–102. New York: Palgrave-Macmillan, 2011.

Dagger, Richard, and David Lefkowitz. 'Political Obligation'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Stanford University, spring 2021. https://plato.stanford.edu/archives/spr2021/entries/political-obligation/.

Darwall, Stephen. 'Reasons, Motives, and the Demands of Morality: An Introduction'. In *Moral Discourse and Practice: Some Philosophical Approaches*, edited by Stephen Darwall, Allan Gibbard, and Peter Railton, 305–12. Oxford: Oxford University Press, 1997.

DePaul, Michael, and Amelia Hicks. 'A Priorism in Moral Epistemology'. In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. Stanford University, fall 2016. https://plato.stanford.edu/archives/fall2016/entries/moral-epistemology-a-priori/.

Enoch, David. 'Agency, Shmagency: Why Normativity Won't Come from What Is Constitutive of Action'. *The Philosophical Review* 115, no. 2 (2006): 169–98.

———. 'Is General Jurisprudence Interesting?' In *Dimensions of Normativity: New Essays on Metaethics and Jurisprudence*, edited by David Plunkett, Scott J. Shapiro, and Kevin Toh, 65–86. Oxford: Oxford University Press, 2019.

———. 'Non-Naturalistic Realism in Metaethics'. In *The Routledge Handbook of Metaethics*, edited by Tristram McPherson and David Plunkett, 29–42. New York: Routledge, 2017. https://doi.org/10.4324/9781315213217-2.

———. *Taking Morality Seriously: A Defense of Robust Realism*. Oxford: Oxford University Press, 2013.

Eriksson, Björn, and Jonas Olson. 'Moral Practice after Error Theory: Negotiationism'. In *The End of Morality: Taking Moral Abolitionism Seriously*, edited by Richard Garner and Richard Joyce, 113–30. New York: Routledge, 2019.

Finlay, Stephen. *Confusion of Tongues: A Theory of Normative Language*. Oxford: Oxford University Press, 2014.

———. 'Defining Normativity'. In *Dimensions of Normativity: New Essays on Metaethics and Jurisprudence*, edited by David Plunkett, Scott J. Shapiro, and Kevin Toh, 187–220. Oxford: Oxford University Press, 2019. https://doi.org/10.1093/oso/9780190640408.003.0009.

———. 'The Error in the Error Theory'. *Australasian Journal of Philosophy* 86, no. 3 (2008): 347–69. https://doi.org/10.1080/00048400802001921.

Foot, Philippa. 'Moral Beliefs'. In *Virtues and Vices and Other Essays in Moral Philosophy*, 110–31. Oxford: Oxford University Press, 2002.

———. 'Morality as a System of Hypothetical Imperatives'. *The Philosophical Review* 81, no. 3 (1972): 305–16. https://doi.org/10.2307/2184328.

———. 'Morality as a System of Hypothetical Imperatives'. In *Virtues and Vices and Other Essays in Moral Philosophy*, 157–73. Oxford: Oxford University Press, 2002.

———. 'Reasons for Action and Desires'. In *Virtues and Vices and Other Essays in Moral Philosophy*, 148–56. Oxford: Oxford University Press, 2002.

Garner, Richard. 'Abolishing Morality'. In *A World Without Values: Essays on John Mackie's Moral Error Theory*, edited by Richard Joyce and Simon Kirchin, 217–33. Dordrecht: Springer, 2010. https://doi.org/10.1007/978-90-481-3339-0_13.

Garner, Richard, and Richard Joyce, eds. *The End of Morality: Taking Moral Abolitionism Seriously*. New York: Routledge, 2019.

Gauthier, David. *Morals by Agreement*. Oxford: Clarendon Press, 1986.

———. 'Why Contractarianism?' In *Contractarianism and Rational Choice: Essays on David Gauthier's Morals by Agreement*, edited by Peter Vallentyne, 15–30. New York: Cambridge University Press, 1991.

Gerritsen, Eline. 'Morality as a Fiction or Agreement: Joyce versus Gauthier'. MLitt dissertation, University of St Andrews & University of Stirling, 2017.

Gibbard, Allan. 'Morality as Consistency in Living: Korsgaard's Kantian Lectures'. *Ethics* 110, no. 1 (1999): 140–64. https://doi.org/10.1086/233207.

———. *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Cambridge, Mass.: Harvard University Press, 1990.

Glüer, Kathrin. 'Sense and Prescriptivity'. *Acta Analytica* 14, no. 23 (1999): 111–28.

Glüer, Kathrin, and Åsa Wikforss. 'Against Content Normativity'. *Mind* 118, no. 469 (2009): 31–70. https://doi.org/10.1093/mind/fzn154.

Goldsworthy, Jeffrey. 'Externalism, Internalism and Moral Scepticism'. *Australasian Journal of Philosophy* 70, no. 1 (1992): 40–60. https://doi.org/10.1080/00048408112340033.

Hampton, Jean. *The Authority of Reason*. Cambridge: Cambridge University Press, 1998.

———. 'Two Faces of Contractarian Thought'. In *Contractarianism and Rational Choice: Essays on David Gauthier's Morals by Agreement*, edited by Peter Vallentyne, 31–55. New York: Cambridge University Press, 1991.

Hattiangadi, Anandi. 'Is Meaning Normative?' *Mind & Language* 21, no. 2 (2006): 220–40. https://doi.org/10.1111/j.0268-1064.2006.00312.x.

———. *Oughts and Thoughts*. Oxford: Oxford University Press, 2007. https://doi.org/10.1093/acprof:oso/9780199219025.001.0001.

Hoffman, Martin L. *Empathy and Moral Development: Implications for Caring and Justice*. Cambridge: Cambridge University Press, 2000.

Hume, David. *A Treatise of Human Nature: A Critical Edition.* Edited by David Fate Norton and Mary J Norton. Oxford: Clarendon Press, 2014.

Hurka, Thomas. 'Underivative Duty: Prichard on Moral Obligation'. *Social Philosophy and Policy* 27, no. 02 (2010): 111–34. https://doi.org/10.1017/S0265052509990173.

Husi, Stan. 'Why Reasons Skepticism Is Not Self-Defeating'. *European Journal of Philosophy* 21, no. 3 (2013): 424–49. https://doi.org/10.1111/j.1468-0378.2011.00454.x.

Jaquet, François. 'Utilitarianism for the Error Theorist'. *The Journal of Ethics* 25, no. 1 (2021): 39–55. https://doi.org/10.1007/s10892-020-09339-x.

Joyce, Richard. 'Moral Fictionalism: How to Have Your Cake and Eat It Too'. In *The End of Morality: Taking Moral Abolitionism Seriously*, edited by Richard Garner and Richard Joyce, 150–65. New York: Routledge, 2019.

———. 'Morality, Schmorality'. In *Morality and Self-Interest*, edited by Paul Bloomfield, 51–76. Oxford: Oxford University Press, 2007. https://doi.org/10.1093/acprof:oso/9780195305845.001.0001.

———. 'Reply to "On the Validity of a Simple Argument for Moral Error Theory"'. *International Journal of Philosophical Studies* 24, no. 4 (2016): 518–22. https://doi.org/10.1080/09672559.2016.1203980.

———. 'The Error In "The Error In The Error Theory"'. *Australasian Journal of Philosophy* 89, no. 3 (2011): 519–34. https://doi.org/10.1080/00048402.2010.484465.

———. *The Evolution of Morality*. Cambridge, Mass.: The MIT Press, 2007.

———. 'The Fugitive Thought'. *The Journal of Value Inquiry* 34, no. 4 (2000): 463–78. https://doi.org/10.1023/A:1004785213821.

———. *The Myth of Morality*. Cambridge: Cambridge University Press, 2001.

Kaplan, Jeffrey. 'The Problem with Descriptive Correctness'. *Ratio* 33, no. 2 (2020): 79–86. https://doi.org/10.1111/rati.12257.

Kiesewetter, Benjamin. 'Are Epistemic Reasons Normative?' *Noûs*, 2021, 1–26. https://doi.org/10.1111/nous.12379.

Kirchin, Simon. 'A Tension in the Moral Error Theory'. In *A World Without Values: Essays on John Mackie's Moral Error Theory*, edited by Richard Joyce and Simon Kirchin, 167–82. Dordrecht: Springer, 2010. https://doi.org/10.1007/978-90-481-3339-0_10.

Kolodny, Niko. 'Why Be Rational?' *Mind* 114, no. 455 (2005): 509–63.

Korsgaard, Christine. *The Sources of Normativity*. Cambridge: Cambridge University Press, 1996.

Krebs, Dennis. 'Empathy and Altruism'. *Journal of Personality and Social Psychology* 32, no. 6 (1975): 1134–46. https://doi.org/10.1037/0022-3514.32.6.1134.

Letsas, George. 'How to Argue for Law's Full-Blooded Normativity'. In *Dimensions of Normativity: New Essays on Metaethics and Jurisprudence*, edited by David Plunkett, Scott J. Shapiro, and Kevin Toh, 165–86. Oxford: Oxford University Press, 2019.

Lord, Errol, and Kurt Sylvan. 'Reasons: Wrong, Right, Normative, Fundamental'. *Journal of Ethics and Social Philosophy* 15, no. 1 (2019): 43–74.

Lutz, Matt. 'The "Now What" Problem for Error Theory'. *Philosophical Studies* 171, no. 2 (2014): 351–71. https://doi.org/10.1007/s11098-013-0275-7.

Mackie, John L. *Ethics: Inventing Right and Wrong*. London: Penguin Books, 1977.

Maguire, Barry, and Jack Woods. 'The Game of Belief'. *The Philosophical Review* 129, no. 2 (2020): 211–49. https://doi.org/10.1215/00318108-8012843.

Manne, Kate. 'Democratizing Humeanism'. In *Weighing Reasons*, edited by Errol Lord and Barry Maguire, 123–40. Oxford: Oxford University Press, 2016. https://doi.org/10.1093/acprof:oso/9780199315192.003.0006.

———. 'Locating Morality: Moral Imperatives as Bodily Imperatives'. In *Oxford Studies in Metaethics 12*, edited by Russ Shafer-Landau, 1–26. Oxford: Oxford University Press, 2017.

———. 'On Being Social in Metaethics'. In *Oxford Studies in Metaethics 8*, edited by Russ Shafer-Landau, 50–73. Oxford: Oxford University Press, 2013.

Mantel, Susanne. 'Do Epistemic Reasons Bear on the Ought Simpliciter?' *Philosophical Issues* 29, no. 1 (2019): 214–27. https://doi.org/10.1111/phis.12149.

McPherson, Tristram. 'Against Quietist Normative Realism'. *Philosophical Studies* 154, no. 2 (2011): 223–40.

———. 'Authoritatively Normative Concepts'. In *Oxford Studies in Metaethics 13*, edited by Russ Shafer-Landau, 253–77. Oxford: Oxford University Press, 2018. https://doi.org/10.1093/oso/9780198823841.003.0012.

Nietzsche, Friedrich. *Beyond Good & Evil: Prelude to a Philosophy of the Future*. Translated by Walter Kaufmann. New York: Vintage Books, 1989.

Nolan, Daniel, Greg Restall, and Caroline West. 'Moral Fictionalism versus the Rest'. *Australasian Journal of Philosophy* 83, no. 3 (2005): 307–30. https://doi.org/10.1080/00048400500191917.

Nussbaum, Martha C. *Frontiers of Justice: Disability, Nationality, Species Membership*. Cambridge, Mass.: Belknap Press, 2007.

Olson, Jonas. 'Error Theory in Metaethics'. In *The Routledge Handbook of Metaethics*, edited by Tristram McPherson and David Plunkett, 58–71. New York: Routledge, 2017. https://doi.org/10.4324/9781315213217-4.

———. 'Getting Real about Moral Fictionalism'. In *Oxford Studies in Metaethics 6*, edited by Russ Shafer-Landau, 181–204. Oxford: Oxford University Press, 2011.

———. 'In Defense of Moral Error Theory'. In *New Waves in Metaethics*, edited by Michael Brady, 62–84. New York: Palgrave-Macmillan, 2011.

———. *Moral Error Theory: History, Critique, Defence*. Oxford: Oxford University Press, 2014.

Parfit, Derek. *On What Matters: Volume One*. Oxford: Oxford University Press, 2011.

———. *On What Matters: Volume Two*. Oxford: Oxford University Press, 2011.

Pigden, Charles R. 'Logic and the Autonomy of Ethics'. *Australasian Journal of Philosophy* 67, no. 2 (1989): 127–51. https://doi.org/10.1080/00048408912343731.

Plato. *Republic*. Translated by C. D. C. Reeve. Indianapolis: Hackett Publishing Company, 2004.

Plunkett, David. 'Robust Normativity, Morality, and Legal Positivism'. In *Dimensions of Normativity: New Essays on Metaethics and Jurisprudence*, edited by David Plunkett, Scott J. Shapiro, and Kevin Toh, 105–36. Oxford: Oxford University Press, 2019.

Plunkett, David, and Scott Shapiro. 'Law, Morality, and Everything Else: General Jurisprudence as a Branch of Metanormative Inquiry'. *Ethics* 128, no. 1 (2017): 37–68. https://doi.org/10.1086/692941.

Prichard, H. A. 'Manuscript on Morals'. In *Moral Writings*, edited by Jim MacAdam, 114–62. Oxford: Oxford University Press, 2002. https://doi.org/10.1093/0199250197.003.0008.

Prichard, H.A. 'Does Moral Philosophy Rest on a Mistake?' *Mind* 21, no. 81 (1912): 21–37.

Prinz, Jesse. 'Against Empathy'. *The Southern Journal of Philosophy* 49 (2011): 214–33. https://doi.org/10.1111/j.2041-6962.2011.00069.x.

Radzik, Linda. 'A Coherentist Theory of Normative Authority'. *The Journal of Ethics* 6, no. 1 (2002): 21–42. https://doi.org/10.1023/A:1015887107235.

———. 'A Normative Regress Problem'. *American Philosophical Quarterly* 36, no. 1 (1999): 35–47.

———. 'Incorrigible Norms: Foundationalist Theories of Normative Authority'. *The Southern Journal of Philosophy* 38, no. 4 (2000): 633–49. https://doi.org/10.1111/j.2041-6962.2000.tb00919.x.

———. 'Justification and the Authority of Norms'. *The Journal of Value Inquiry* 34 (2000): 451–61.

Rawls, John. *A Theory of Justice*. Cambridge, Mass.: Belknap Press, 1971.

Sayre-McCord, Geoffrey. 'Deception and Reasons to Be Moral'. In *Contractarianism and Rational Choice: Essays on David Gauthier's Morals by Agreement*, edited by Peter Vallentyne, 181–95. New York: Cambridge University Press, 1991.

Scanlon, Thomas. *What We Owe to Each Other*. Cambridge, Mass.: Belknap Press, 2000.

Searle, John R. 'How to Derive "Ought" From "Is"'. *The Philosophical Review* 73, no. 1 (1964): 43. https://doi.org/10.2307/2183201.

Shafer-Landau, Russ. 'Error Theory and the Possibility of Normative Ethics'. *Philosophical Issues* 15, no. 1 (2005): 107–20. https://doi.org/10.1111/j.1533-6077.2005.00056.x.

———. *Moral Realism: A Defence*. New York: Oxford University Press, 2005.

Silverstein, Matthew. 'Inescapability and Normativity'. *Journal of Ethics and Social Philosophy* 6, no. 3 (2012): 1–27. https://doi.org/10.26556/jesp.v6i3.67.

Sinnott-Armstrong, Walter. *Moral Skepticisms*. Oxford: Oxford University Press, 2006.

Smith, Holly. 'Deriving Morality from Rationality'. In *Contractarianism and Rational Choice: Essays on David Gauthier's Morals by Agreement*, edited by Peter Vallentyne, 229–53. New York: Cambridge University Press, 1991.

Smith, Michael. *The Moral Problem*. Oxford: Blackwell, 1995.

Sobel, David. *From Valuing to Value: A Defense of Subjectivism*. Oxford University Press, 2016.

Southwood, Nicholas. 'Vindicating the Normativity of Rationality'. *Ethics* 119, no. 1 (2008): 9–30. https://doi.org/10.1086/592586.

Stohr, Karen. *On Manners*. New York: Routledge, 2011.

Street, Sharon. 'A Darwinian Dilemma for Realist Theories of Value'. *Philosophical Studies* 127, no. 1 (2006): 109–66. https://doi.org/10.1007/s11098-005-1726-6.

———. 'Coming to Terms with Contingency: Humean Constructivism about Practical Reason'. In *Constructivism in Practical Philosophy*, edited by James Lenman and Yonatan Shemmer, 40–59. Oxford: Oxford University Press, 2012. https://doi.org/10.1093/acprof:oso/9780199609833.003.0003.

———. 'Constructivism About Reasons'. In *Oxford Studies in Metaethics 3*, edited by Russ Shafer-Landau, 207–45. Oxford: Oxford University Press, 2008.

———. 'In Defense of Future Tuesday Indifference: Ideally Coherent Eccentrics and the Contingency of What Matters'. *Philosophical Issues* 19 (2009): 273–98.

———. 'Nothing "Really" Matters, but That's Not What Matters'. In *Does Anything Really Matter?: Essays on Parfit on Objectivity*, edited by Peter Singer, 121–48. Oxford: Oxford University Press, 2017.

———. 'What Is Constructivism in Ethics and Metaethics?' *Philosophy Compass* 5, no. 5 (2010): 363–84. https://doi.org/10.1111/j.1747-9991.2009.00280.x.

Streumer, Bart. *Unbelievable Errors: An Error Theory about All Normative Judgements*. Oxford: Oxford University Press, 2017.

Tiffany, Evan. 'Deflationary Normative Pluralism'. *Canadian Journal of Philosophy* 37, no. 5 (2007): 231–62. https://doi.org/10.1353/cjp.0.0076.

Tiffany, Evan, and Sam Black. 'Moral Philosophy Does Not Rest on a Mistake: Reasons to Be Moral Revisited'. In *Reasons to Be Moral Revisited*, edited by Evan Tiffany and Sam Black, vii–xl. Canadian Journal of Philosophy Supplementary, v. 33. Calgary: University of Calgary Press, 2007.

Werner, Preston J. 'Why Conceptual Competence Won't Help the Non-Naturalist Epistemologist'. *Canadian Journal of Philosophy* 48, no. 3–4 (2018): 616–37. https://doi.org/10.1080/00455091.2017.1410417.

Wiggins, David. *Ethics: Twelve Lectures on the Philosophy of Morality*. London: Penguin Books, 2006.

Williams, Bernard. 'Internal and External Reasons'. In *Moral Luck: Philosophical Papers, 1973-1980*, 101–13. Cambridge: Cambridge University Press, 1981.

Wodak, Daniel. 'Mere Formalities: Fictional Normativity and Normative Authority'. *Canadian Journal of Philosophy* 49, no. 6 (2019): 828–50. https://doi.org/10.1080/00455091.2018.1433795.

Woods, Jack. 'The Authority of Formality'. In *Oxford Studies in Metaethics 13*, edited by Russ Shafer-Landau, 207–29. Oxford: Oxford University Press, 2018. https://doi.org/10.1093/oso/9780198823841.003.0010.