

Politics, Philosophy
& Religion

Lancaster
University



**Defending Normative Realism Against Evolutionary Debunking
Arguments**

By

Ryan Anderson Bartle, BA, MA

Supervised by Dr Christopher Macleod and Dr Brian Garvey

**A thesis submitted in partial fulfilment of the requirement for the degree
of:**

PhD in Philosophy

25/09/2022

Declarations

Word Count: 79,962

I can confirm that the above word count does not exceed the permitted maximum.

I declare that the thesis is my own work, and has not been submitted in substantially the same form for the award of a higher degree elsewhere.

Ryan Anderson Bartle (25/09/2022)

Contents

Acknowledgements	6
Abstract	7
Introduction	8
Chapter 1: A Statement of Normative Realism	17
1. Normative Realism	17
2. Naturalistic Realism	23
2.1. The Case Against Analytic Naturalism	25
2.2. The Case Against Non-Analytic Naturalism	27
2.3. Non-Naturalistic Realism	29
2.4. Normative Rationalism	30
2.5. Ethics Apart From Metaethics?	35
3. Objections	37
3.1. The Argument from Physicalism/Naturalism	37
3.2. The Charge of Platonism	38
Chapter 2: Evolutionary Debunking Arguments	43
1. Evolution and Normativity	44
1.1. The Evolutionary Process	44
1.2. The Relation Between Evolution and Normativity	46
1.3. The Evolution of Norms: The Empirical Evidence	51
2. Evolutionary Ethics	57
2.1. An Evolutionary Justification For Normativity	57
2.2. Evolutionary Debunking Arguments	63
2.3. The Distinction Between Content Partial and Content Neutral EDA	71
Chapter 3: Has Evolution Shaped Our Normative Judgments?	76
1. Previous Realist Responses to the Darwinian Dilemma	76
1.1. The Argument from First-Order Normative Truth	76
1.2. Resisting the Massive Coincidence Objection	79
2. More Promising Lines of Response	80
2.1. Does Anti-Realism Follow Logically from Street's Argument	80
2.2. The Argument from Non-Adaptive Normative Beliefs	85

2.3. An Objection: Is This Argument a Red Herring?	98
3. An Argument from Disagreement	99
3.1. Explaining Moral Disagreement	104
3.2. Intercultural and Intracultural Moral Disagreement	105
3.3. Intergenerational Disagreement	106
3.4. Intrapersonal Disagreement	106
4. Objection: Does This Argument Miss the Target?.....	110

Chapter 4: Can Evolution Shape Normative Judgment Without Undermining it?114

1. Evolutionary Relations	114
1.1. Searching for the Problematic Relation	114
1.2. The Tracking Account of the Relation Between Evolution and Normative Truth ..	116
1.3. Causing, Shaping, and the Darwinian Dilemma Revisited	120
2. Evolution and Hedonistic Disvalue	123
2.1. Realism and Pain’s Badness	124
2.2. Why is Pain Bad?	124
3. The Evolutionary Debunking Argument Against Hedonism	131
3.1. How Evolution Usually Shapes Normative Judgments: Haidt’s Social Intuitionist Model	133
3.2. Why the Social Intuitionist Model Does Not Explain Pain’s Badness	135
4. Objections	139
4.1. Same Debunking, Different Intermediary	139
4.2. The Contingency of Pain’s Badness	140

Chapter 5: Where Does Normative Knowledge Come From?144

1. Normative Genealogies	145
1.1. An Outline of a Dualist Normative Genealogy	146
1.2. The Role of Evolution	146
1.3. Where Joyce’s Genealogy Falls Short	148
1.4. Implications For Realism Vs Anti-Realism	153
2. Normative Epistemology	154
2.1. The Structure of Normative Knowledge	154
2.2. Problems with Reflective Equilibrium	155
2.3. Foundationalism and Intuitionism	157
2.4. Toward a Founherentist Normative Epistemology	166
3. Dealing with Distortion	168
3.1. Sinnott-Armstrong’s Empirical Case Against Intuitionism	168
3.2. Intuitions and Evolution	170
3.3. An Outline of a Normative Methodology	173

Concluding Remarks182

Bibliography 187

Acknowledgements

I am very grateful for the assistance provided by my two supervisors, Dr Christopher Macleod and Dr Brian Garvey. Both of whom provided valuable advice and constructive criticism of the draft chapters they read and were very generous with their time. I would like to thank Dr Neil Manson and Professor Chakravarthi Ram-Prasad for their observations and suggestions during my appraisal meetings, which were always insightful and useful.

I owe thanks to the Arts and Humanities Research Council (AHRC), which supported my research and without which, the work carried out here would not have been possible.

I would like to thank my family for all the support they have given me throughout my time as a PhD student, above all my wife Abigail, my mother Sharon and step mother Brigid, my Dad Thomas and my Grandparents Charles and Margaret.

Defending Normative Realism Against Evolutionary Debunking Arguments

Ryan Anderson Bartle BA, MA

A thesis submitted in partial fulfilment of the requirement for the degree of PhD in Philosophy

Abstract

Normative realism is a view characterised by the belief in stance-independent normative facts. The primary aim of this thesis is to defend normative realism against Evolutionary Debunking Arguments (EDA). EDA are used in several philosophical disciplines, including metaethics, normative ethics and epistemology. Because I am defending a metaethical position, my interest is with their use in metaethics. Perhaps the most well-known EDA is Sharon Street's *Darwinian Dilemma for Realist Theories of Value* (2006). The problem for realists, Street thinks, is that if we accept their claim that stance-independent normative facts exist, the fact evolution has "thoroughly saturated" our normative judgments means we are unlikely to ever know them (2006, pp.113-114). This is because evolution would only shape our normative judgments toward truth if normative knowledge promoted genetic survival and reproduction, an idea Street rejects. Another highly influential EDA is outlined by Richard Joyce in *The Evolution of Morality* (2006). Joyce likens morality to a useful fiction. We evolved to believe in morality not because it has any basis in fact, but because it motivated our species to act in pro-social ways that contributed to survival and reproduction. Chapter 1 outlines the non-naturalistic realist view that I aim to defend against EDA. Chapter 2 outlines the major premises and conclusions of EDA. Chapters 3 and 4 defend realism against Street's EDA – in the process of which, I show, among other things, that realism combined with hedonistic utilitarianism is immune to EDA. Chapter 5 responds to Joyce's EDA and addresses normative epistemological and methodological questions raised by the prospect that many of our normative beliefs are distorted by evolution. Here I draw upon three leading theories of epistemology (*Foundationalism, Coherentism, Reliabilism*) in devising a multi-layered methodology for the purposes of *justifying* normative principles and guarding against potential sources of distortion.

Introduction

Normative realism is a metaethical theory about the nature and epistemic status of normative judgments. According to the realist, the surface grammar of normative claims is not misleading. Normative judgments like *murder is wrong* and *abortion is permissible* look like assertions of fact because they are (Shafer-Landau 2003, pp.17-18). Not all realists agree on the content of normative truths. Realists can be virtue ethicists, utilitarians, divine-commandment theorists, or proponents of any other normative theory; but what they all agree on, is that normative facts exist (ibid, pp.13-14). Realists also hold that normative facts are stance-independent. If murder is wrong, it is not simply because I personally disagree with it. Murder is wrong regardless of what I, my community, or anybody else, happens to think about murder.

Some normative judgements are about *morality*. *Murder is wrong* and *abortion is permissible* seem to fall within that category, though precisely what makes a claim moral, as opposed to some other species of normativity is a contested matter. Philosophers may agree that some x is a normative truth but disagree about whether it is a specifically moral truth. For example, Roger Crisp holds a broadly act utilitarian view (Crisp 2006, ch.1), but considers this to be a normative theory as opposed to a moral theory. Likewise, it is unclear whether rational egoism, the theory that we ought to maximise our own individual welfare, is best conceived as a moral theory or a theory of rationality. Though I consider myself to be a moral as well as a normative realist, the aim of this thesis is to defend normative realism. Therefore, I will leave aside the question of how we distinguish between different kinds of normative judgments. My intention is to show that a particular version of normative realism is a viable metaethical position and that it can be defended against a range of anti-realist arguments.

What specifically makes a judgment normative? Typically, normative claims are distinguished from descriptive claims in being about what *ought* to be the case, as opposed to what *is* the case. This is a good starting point. That the table in front of me is brown is clearly a descriptive claim. That *I ought to believe* the table is brown is a normative claim about what I have epistemic reason to believe. Moral claims are clearly normative in that they tell us *what we ought to do*. The same is also true of claims about prudential rationality, such as *you ought not to prioritise the future over the present* and *you ought not to smoke because it is bad for you*. Normative claims are not universally *ought* claims. Some of the above can be rephrased to omit

the word ought yet remain normative. E.g., *the correct action is to not smoke because it is bad for you*. Some claims about value possess descriptive surface grammar too. This is generally true of claims about aesthetics and wellbeing, such as *the Mona Lisa is beautiful* and *looking at beautiful things is good for you*. Despite their descriptive quality, these claims seem on reflection to be normative.

I doubt it is possible to lay out necessary and sufficient conditions which cover all normative claims. The suggestion that they involve *oughts* is at best a sufficient but not necessary condition. What many normative claims appear to have in common is that they explicitly or implicitly suggest reasons for action. To claim that maximising happiness *is right, what I ought to do, or my duty* is to claim that I have some reason to maximise happiness. There may be disagreement about whether this is my ultimate reason or whether this reason can be defeated by side constraints. But if I am claiming that we ought to maximise happiness, I must believe we have some reason to do so. Value claims are less straightforwardly action-guiding. There is nothing logically contradictory about claiming that *happiness is good for us but we have no reason to be happy*. Nonetheless, such a claim seems highly counterintuitive. If happiness is good for us, it is natural – and I would suggest rational – to hold that we have some reason to want to be happy, again with the caveat that this may neither be our ultimate reason for action, nor infeasible by other considerations.

In this thesis, I will defend the claim that some normative claims are true. This is my ultimate aim, though I do have other secondary aims. Normative realism is characterised by belief in stance-independent normative facts, but normative realists themselves typically take a stance on what those *first-order* truths are. The first-order view I accept is a version of hedonistic utilitarianism and defending this theory will be one of the secondary aims of this thesis. Hedonistic utilitarianism is the view that the right action is to impartially maximise valuable mental states.¹ That I am defending a first-order normative claim within a thesis which is primarily about metaethics will raise some eyebrows, especially among those who hold that meta and normative ethics are separate spheres of inquiry that should proceed independently from one another. This criticism is sound to the extent that metaethical views do not imply any particular set of first-order commitments. Utilitarians can be cognitivists, non-cognitivists, realists or anti-realists, just as realists can be utilitarian or anti-utilitarian. But it does not thereby follow that the two spheres of inquiry are wholly independent. I will argue in this thesis

¹ I refrain from using the term *pleasure* as that to be maximised for reasons outlined in chapter four.

that whether or not a realist's view is immune to certain anti-realist arguments will depend in part on what first-order normative beliefs she holds.

Over the course of the next five chapters, I aim to defend realism against a range of views. But my primary contribution will be to provide a comprehensive defence of realism against so-called *Evolutionary Debunking Arguments* (EDA). Though these arguments come in different forms, a common-thread linking metaethical EDA² is the claim that the evolutionary origins of our normative judgments undermine the realist's claim that such judgments can be stance-independently true. Proponents of these arguments, notably Sharon Street (2006) and Richard Joyce (2006), appeal to insights in evolutionary psychology and primatology that provide evidence that our normative thinking and practices evolved. Street points out that being able to recognise normative truths is unlikely to provide an organism with an evolutionary advantage (Street 2006, pp.130-131). In which case, even if we assume there are normative truths, evolution's influence on our normative beliefs is unlikely to have been in a truth-tracking direction. This at least implies that we should be sceptical about the realist's claim that we can know normative facts. According to Joyce, the very belief in the existence of normative facts was basically a useful adaptation which provided our species with the psychological and motivational impetus to transcend our naked self-interest and act in ways which promote social cohesion and cooperation (Joyce 2001, pp.138-140). Morality, in this respect, is akin to a "myth,"³ albeit one instrumental to the flourishing of highly social species like humans.

I will be defending normative realism against EDA. However, I believe that any such defence will need to be a nuanced one, given the clear force of evolutionary approaches to ethics. Though I think proponents of EDA often exaggerate the extent to which evolution shaped our normative thinking, it is hardly plausible to hold that it holds no explanatory power given that humans and their mental faculties are products of natural selection. An adequate defence of realism will somehow need to reconcile the likely fact that many of our normative beliefs are to a large extent shaped by what is basically an amoral process, with their belief that we can recognise normative truths. This is my primary objective. I will now discuss the steps by which I hope to establish this conclusion.

In chapter one, I provide a statement of normative realism and respond to some of the general objections this theory faces. The need for specificity in outlining my own realist position is

² Some EDA, as I note in chapter 2, are directed toward normative as opposed to metaethical views.

³ Hence the title of Joyce's book: *They Myth of Morality* (2001).

necessitated by first, the fact that not all versions of realism are vulnerable to the same objections and second, because in the final chapter, I will be providing an account of how we recognise normative facts and this requires my taking a position on the nature of normative facts. Realist views are generally divided into naturalist and non-naturalist camps. Both camps agree that normative facts exist, but disagree metaphysically about what kind of facts they are, and consequently epistemologically about how we recognise them. Naturalistic realists hold that normative facts are basically a species of natural fact. Consequently, we come to recognise them by the same process that we recognise other natural facts: *sense perception*. In this respect, naturalists tend to be empiricists about how we acquire normative knowledge. An alternative view, and the one I accept, is non-naturalism. According to the non-naturalist, it is a category error to claim that normative facts are natural. Concepts such as *ought*, *right*, *good* and *bad* do not form part of the fabric of the natural physical world and consequently cannot be recognised solely by empirical means. Of course, the object of our normative judgments is often part of the natural world. The assertion that *murder is wrong* clearly denotes a collection of natural events that we define as characterising the act of murder. But the statement that murder is wrong is not an analytic one. Wrongness is not the same thing as murder (other things are wrong after all), but a property of it. Though I can *observe* someone being murdered, I cannot observe *murder's wrongness*, because wrongness is an abstract non-natural property which supervenes on certain acts, such as murder.

Since I deny that normative facts can be recognised empirically the only other option is that they are recognised *a priori*. That is, by the exercise of our ability to reason, as opposed to sense perception. From the early 20th century, the idea that we can recognise non-analytic truths a priori came under increasing criticism from various quarters. Logical positivists like A.J. Ayer were perfectly happy to afford a role to the a priori, but only in relation to analytic statements (Ayer 1952, ch.4). Hard naturalists and physicalists argued against the metaphysical commitments of the non-naturalist by denying non-natural truths a place within their ontology. Moreover, the general metaethical climate around this time was largely hostile to realist views. The brief dominance of non-cognitivist views called into question the very idea that normative claims are assertions of truth.⁴ And the 'linguistic turn' shifted the focus of ethics significantly from an exercise of discovering universal truths to the study of the nature of ethical language. Though many of the arguments against realism have considerable force, I argue that they are surmountable. Though many philosophers are sceptical about the existence of non-natural

⁴ E.g. Ayer (1952, ch.6) and Hare (1991, pp.451-462).

truths, some truths are difficult to conceptualise in a purely natural or physical way. Mathematical truths like $2+3=5$ and logical truths like *nothing can be both true and false at the same time*, are not observable entities like tables and chairs. Insofar as these propositions are true it is hardly because they form part of the fabric of the natural world. Moreover, no inductive argument can inform us of these truths and they cannot be deduced from prior premises. I would suggest the only promising account of how we know them must give significant weight to the a priori. If this is true, then we have already opened up our ontology to include non-naturalistic truths.

Having outlined my metaethical commitments in the first chapter, I then go on to outline the arguments which I will be defending this view against: *Evolutionary Debunking Arguments*. As a gateway into this discussion, I begin by exploring the historical linkage between evolutionary theory and ethics. Though the use of EDA is a recent phenomenon in academic philosophy, they have antecedents going back to the intellectual climate in the early aftermath of the publication of Darwin's theory of natural selection. For example, Richard Joyce claims that evolution undermines belief in normative truth, and that as such, a proper understanding of the role of evolution in shaping human judgments should lead us toward normative scepticism (Joyce 2013, p.351). It is worth noting that this claim bears some similarity to the claims of early Christians who were concerned that an acceptance of Darwin's theory undermines morality.⁵ Some of these concerns emanated from misunderstandings of the theory, for which some of those who called themselves Darwinians deserve some share of the blame. An example being how some overemphasised *selfishness* and *struggle* over other factors in the selective process, such as *altruism* and *reciprocity*, giving the false impression that any evolved morality would necessarily be cold-hearted and cruel.⁶ But a more pressing concern has stood the test of time. That if humans and their intellectual faculties are products of natural selection - an amoral process - how could it be that we were able to recognise normative truths?

Having discussed some of these historical debates, I outline some contemporary work in primatology and evolutionary psychology that seeks to identify evolutionary origins for our normative judgments. I then go on to outline how this work has formed the basis of EDA. EDA are not restricted to metaethics. So-called *local* EDA are used in normative ethics to undermine individual normative principles by arguing that they originated in evolution. These arguments have no obvious metaethical significance because even if they are successful, they would only

⁵ E.g. Wilberforce (1860).

⁶ I provide some examples in the first chapter.

undermine the truth or justification of the principle being considered, not the general claim that some normative beliefs are true. Hence, local EDA are not my primary focus – though they are relevant to my secondary aim of defending a particular first-order view. My primary focus is on *global* EDA. These arguments are of metaethical significance since they take aim at a metaethical theory: realism.⁷ How they do this varies. I draw an original distinction between what I call *content-partial* and *content-neutral* EDA. Content partial EDA attempt to undermine realism by showing that the content of most of our first-order normative beliefs was shaped by evolution.⁸ Assuming that evolution has a distortive influence on our normative judgments, it would follow that most of our normative beliefs are off-track. Even if normative truths exist, under these epistemic conditions – ones in which our normative beliefs are predominantly off-track – the likelihood that we would ever come across them is slim. Joyce’s argument is what I call *content-neutral* in that he remains neutral about whether evolution shaped the content of our normative beliefs (Joyce 2006, pp.180-181). His argument is that evolution is responsible for the fact that we think in a normative way in the first place and our *meta* belief that there are such things as normative facts which are capable of binding us irrespective of our personal preferences. Joyce thinks that morality is basically deserving of the status of myth. It was a useful myth insofar as it provided the psychological motivation for behaviour that transcends self-interest for the good of society (and thereby indirectly good for individuals) but it has no basis in fact. In the subsequent three chapters, I defend normative realism against both versions of the EDA.

Chapters three and four consist of responses to Street’s content partial EDA. In chapter three, I explore a range of responses to her claim that evolution shaped most of our normative beliefs. I note that even if this claim is true, it still leaves open the possibility that a number – albeit perhaps a small number – of our normative beliefs were shaped by other factors. Not only is this a conceptual possibility, but some normative beliefs seem to fall within that category. For instance, if beliefs such as *it is right to give priority to kin over strangers* can be explained by evolution, then a belief such as utilitarian impartiality, which claims that we should be impartial in how we distribute benefits and burdens, seems to fall a-foul of an evolutionary explanation. I suggest that several utilitarian beliefs are of this kind, as well as various other normative beliefs. Insofar as such beliefs cannot be explained by evolution, they cannot be undermined by evolution and are thus immune from Street’s argument. This is one respect in which

⁷ Both Street and Joyce see realism as the target of their arguments. For a similar argument, see Ruse (1995).

⁸ A primary example of this argument can be found in Street (2006).

normative and metaethical views are connected. The argument pursued in this chapter, if successful, would only vindicate normative realists whose first-order normative views cannot be explained by evolution.

Another key argument of this chapter is a revised version of the *argument from disagreement*. The original argument from disagreement was devised by John Mackie as an objection against realism (Mackie 1977, pp.36-38). Mackie claimed that the empirical fact that humans disagree so much about ethics is evidence against there being any objective normative facts. According to my variation of this argument, the fact that humans normatively disagree so much is evidence against the view that most of our normative judgments were caused by evolution. Modern human beings basically share the same evolutionary history, which means that if evolutionary pressures are responsible for our normative beliefs, we should expect normative uniformity. But this has not obtained. There is widespread disagreement. From this I infer that evolution is but one of several factors which influence our normative judgments and by no means the only or most salient one.

In chapter four, I focus on a different aspect of Street's argument, namely her claim that if evolution shapes a normative judgment, it is most likely not in the direction of truth, but in whatever direction happened to aid survival and reproduction. The arguments pursued in chapter three do not undermine this claim. Even if certain principles can be defended by showing that belief in them was not caused by evolution, this would not vindicate normative beliefs that were caused by evolution. As it happens, I think that an evolutionary explanation of a normative principle does tend to undermine that principle. However, I argue that there are counterexamples. One of which is the belief in a hedonistic theory of value. Hedonism has an evolutionary explanation, but this explanation is unique insofar as it does not tend to undermine the theory. It might be thought that this result is very convenient given my secondary aim of defending hedonistic utilitarianism. Perhaps so, but I have an argument for it and convenient or not, my conclusion should stand or fall with the force of that argument.

According to Jonathan Haidt's (2001) influential account of normative judgment, evolution shaped our normative beliefs via our *moral emotions*, such as *disgust*, *sympathy* and *shame*. I argue that normative views about mental states – such as the belief that *pain is bad* and *pleasure is good* – are exceptions to this rule, insofar as such beliefs are not primarily shaped by emotion. My focus is primarily on *pain is bad*. Why do we hold this view? Clearly, evolution is causally relevant to the extent that we evolved to have negative responses to pain phenomena because

it indicates danger to an organism's life – such as the pain of burning, which indicates the presence of fire. But the way we recognise pain as being bad has more to do with the way it feels, as opposed to its being the object of a negative emotional reaction. For instance, picking up a scolding tray is a bad experience, but the pain involved is usually so brief that it does not allow for emotions to come into play until after the pain has ceased. I argue that because our belief in pain's badness is shaped by our conscious experience of pain as opposed to a preconditioned emotional response, this belief is more epistemically secure than those normative beliefs shaped by intermediary emotional states.

In chapter five, I turn my attention to Joyce's content neutral EDA. Again, I think this view contains a hefty dose of truth and my disagreement is a partial one. I accept Joyce's claim that certain aspects of human normativity can be explained by evolution. My objection is that Joyce's normative genealogy is too limited. Normativity is multifaceted. On the one hand there are normative beliefs which reflect instinctive dispositions such as *retributivism* and its manifestation in beliefs such as *an eye for an eye*. On the other hand, there are more reflective normative beliefs which often go against our instincts. For example, it is human nature to hold a grudge against someone who has wronged you and this may lead to an instinctive desire to see them punished. But we can also think about this situation in a more impersonal and less impassioned way. I can critically reflect on whether retributive punishment is a justified practice. And even if I still have a basic desire to see revenge taken on my wrongdoer, I might nonetheless conclude, on the basis of reflection, that retributive punishment is unjustified. This is just to illustrate that there is often a conflict in our normative thinking between our basic instinctive normative beliefs and those which arise out of a process of critical thinking. I argue that instinctive normative beliefs by their very nature are most likely to correspond to Joyce's evolutionary genealogy. Whether this applies to our reflective normative beliefs is much more questionable. These beliefs, I suggest, originate in our reasoning capabilities, which are more difficult to debunk than our evolved instincts and emotions because they have a track record of leading us toward truth.

The arguments made by this point will have shown that evolution's role in shaping human ethics is less considerable than evolutionary debunkers have claimed. Nonetheless, in accepting that many of our normative judgments have been shaped by evolution, I concede that its influence on human ethics has been considerable. This raises certain epistemological and methodological considerations. First, since evolution has influenced our normative judgments in a generally distorting way, how do we go about recognising normative truths? I explore what

answers the traditional theories of epistemology, namely *coherentism* and *foundationalism*, can provide. Ultimately, though both theories have their merits, I find them both deficient in key respects and argue for a *foundherentist* approach to normative epistemology, which gives weight both to coherence and self-evidence.⁹ This naturally leads to methodological questions about whether some decision-making procedure can be constructed which enables us to identify potential sources of distortion, such as evolution, and thereby avoid it. In the second part of chapter five, I will attempt to outline a methodology to achieve this aim.

⁹ The foundherentist approach to epistemology was devised and developed by Susan Haack (1993).

Chapter 1: A Statement of Normative Realism

Introduction

Like other metaethical positions, realism is more of a family of views than a unified theory. This creates complication when it comes to defending realism, for differences in metaphysical and epistemological positions mean that no such universal defence is available. To illustrate, some realists hold strongly naturalistic views concerning the nature of normative facts and empiricist views about how we know such facts.¹⁰ Others claim that normative facts are non-naturalistic and can be known a priori.¹¹ Clearly, such theorists will need to offer different answers to the evolutionary debunker's question of how we might have evolved to know normative truths. Many of the arguments made in this thesis are open to all realists; but this is not always the case. Therefore, rather than being a defence of realism generally, this thesis is primarily a defence of one version of realism. In this chapter, I will outline the realist view I am defending. I begin in section 1 by explaining the fundamental tenets of realism and give some reasons for why it is an attractive metaethical position. In section 2, I provide several arguments against *naturalistic* versions of realism before going on to outline what I believe is a more plausible alternative: *non-naturalistic realism*. In section 3, I offer some responses to some of the most pressing objections against this view (excluding EDA).

1. Normative Realism

Not all normative realists are moral realists; but since morality is a species of normativity, all moral realists are normative realists. But normativity extends beyond morality. It encompasses claims about non-moral goodness and badness, which, for teleological theories like utilitarianism, play a prominent role in moral theory. It also encompasses claims about rationality, whether that be prudential claims about what I have self-interested reason to want or do for my own sake,¹² or epistemic claims about what a rational person ought to believe. It may well encompass a lot more, but my interest is broadly with these kinds of claims. I will be defending the view that some normative claims are true. That in some (perhaps all) situations, there are right and wrong ways to act.

¹⁰ See Copp (2007, ch.1) for a thorough overview of the metaphysical and epistemological commitments of naturalistic realism.

¹¹ For non-naturalist views see: Shafer-Landau (2003, ch.3), Parfit (2011b, ch.31).

¹² For an argument that prudential claims qualify as a species of the normative, see: Fletcher (2019, pp.81-91).

A belief that there are normative truths lies at the heart of the realist's thesis (Shafer-Landau 2003, pp.13-14; Parfit 2011b, pp.464-475). This is an ontological claim about what kind of things exist. Realists are also committed to a particular semantic thesis about how our normative language relates to these truths; namely, *cognitivism* (Shafer-Landau 2003, pp.17-18). This is the view that, when uttered sincerely, normative statements attempt to express facts about how things are, and as such, convey *beliefs*. In making this claim, realists reject the non-cognitivist thesis that normative claims express non-doxastic mental states, such as emotional approval or disapproval, an endorsement of a set of norms, or a universal prescription.¹³ According to the realist, if a person sincerely claims that murder is wrong, they are expressing their belief that it is true that murder is wrong. I use the word "sincerely" here because a realist may accept that normative language is sometimes used in a non-doxastic way. Perhaps in a heated exchange, someone might yell: "you're in the wrong!" when what she really is doing is expressing her anger or disgruntlement at what the other person has said. I think that normative language is sometimes used in this way. But to qualify as a realist, one need not say that normative statements always express beliefs, but that they sometimes do.

We have, then, two necessary conditions for realism. One being that normative statements (sometimes) express beliefs. The other being that - contrary to cognitivist error theorists¹⁴- there are normative facts which can be the object of such beliefs. These claims are not jointly sufficient. Some anti-realist subjectivists and constructivists, such as Sharon Street, believe that there are normative truths (Street 2010, pp.366-367). As do some moral relativists (Suikkanen 2019). What sets these philosophers apart from realists is that they believe all such truths to be in some way contingent upon the evaluative beliefs or value judgments people hold. For subjectivists, an agent's reasons for action come from her own set of values and desires. If, to use Street's example, Caligula desires nothing more than to torture people for fun, then, assuming this desire does not conflict with his other desires, he has a normative reason to torture people for fun (Street 2010, pp.270-371). Since the distaste and disgust with which most of us view torture does not bear upon Caligula's value judgments, this normative claim has no hold upon him. In the case of relativists, for a normative claim to have hold upon an agent in some community (however that is defined) it must resonate with the normative judgments held by the members of that community.

¹³ For non-cognitivist views, see: Hare (1991, pp.451-462); Gibbard (1990); Schroeder (2008).

¹⁴ Cognitivist error theorists hold that normative judgments express beliefs, but that such beliefs are uniformly false; see Mackie's (1977, Ch.1). For an error theory based on EDA, see Joyce (2006, Ch.6).

Realists reject these claims. Some philosophers¹⁵ have cashed this disagreement out in terms of the distinction between mind-dependence and mind-independence. The anti-realist, it is said, holds normative truths to be contingent in some way on the minds of agents, whereas realists hold normative truths to be mind-independent. Like Shaffer-Landau (2003, p.15) I find this way of framing the difference problematical.¹⁶ It seems, for instance, to disqualify the realist from holding certain first order normative beliefs, particularly those relating to mental states. Consider hedonism. As a normative theory, hedonism is presumably compatible with any metaethical position. But if the realist must be committed to the claim that normative truths are mind-independent, then it would seem that hedonism and realism are incompatible, for the mental states to which hedonists ascribe value are clearly not mind-independent. Still, intuitively it seems that if someone claims a) that pleasure is the only good, b) that this claim is true and c) that anyone who denies this claim is wrong, then such a person is a realist. A better way to cash out the difference between realists and anti-realists is in terms of stance-independence. On the anti-realist view, normative truths are stance-dependent insofar as they are contingent on the evaluative or value judgments agents hold. This is the view that realists deny. Consider the hedonistic normative realist again. She claims that the one and only good is pleasure. She also claims that even if an agent does not value, want, or desire pleasure, or in any way judge pleasure to be good, it would nonetheless still be true that *pleasure is good for her*. By committing the realist to stance-independence rather than mind-independence, we leave the option open for the realist to ascribe normative value to mental states whilst still preserving the fundamental realist notion that people can be substantially mistaken about what normative reasons they have.

These three conditions are often claimed to be sufficient for one's qualifying as a realist. In practice, realists seem to be committed to more. It is a possibility, for instance, that due to some kind of cognitive limitation, humans do not possess the ability to discover these stance-independent truths. However, for the most part, realists seem committed to the additional claim that humans are capable of recognising normative truths. There is a degree of ambiguity about

¹⁵ Street (2010, p.370) tells us that the main component of the disagreement between realists and anti-realists is "the question of mind-dependence." Although in explaining why her own constructivism is a brand of anti-realism, she cites its appeal to the "practical point of view". This seems more precise than the term mind-independent, and I think that if we were to say that realists are marked out by claiming that normative truths are independent of any agent's "point of view", such a description would be virtually, if not completely, equivalent to my own.

¹⁶ Shafer-Landau uses the term *stance-independence* to describe the realist account of normative truth. I follow this way of framing the distinction in this thesis, though other ways of framing it could potentially work, e.g. independence of *point of view, opinion, belief, perspective* etc. Some of these terms are close to being, if not are, synonymous and therefore there is no wonder that they can play the same role for the realist.

what, exactly, this means. Presumably the realist need not claim that everyone can know about these normative truths; or for those who can, that they in fact do come to recognise them. But at the same time, she would also not want to claim that no one can know them; or even that knowledge of them is confined to a privileged epistemic elite.¹⁷ This issue has not received much attention from realists and therefore it is better to speak for myself on the matter. Intuitively, normativity seems unlike quantum mechanics or advanced mathematics. Even if there is such a thing as a normative genius, it seems unlikely that one must be a genius in order to navigate the normative essentials. This does not mean that every human does or will in fact grasp normative truths, but they are not, in principle, cognitively out of reach (though, as I will discuss in subsequent chapters, there may be distortive influences which serve to inhibit our access to normative truth).

We have, then, the following four tenets of realism.

A person is a realist if they believe that:

- (1) Cognitivism is true
- (2) There is either a single normative truth or a plurality of normative truths
- (3) Normative truths are not contingent upon agential judgements
- (4) The average human being is capable of recognising normative truths

What reasons are there for accepting realism? As I see it, there are two primary reasons, both of which are very straightforward. The first is that the realist's tenets just seem to cohere with many people's intuitions about the nature of normative claims; the second is that, on the face of it, anti-realist views have counterintuitive implications. Let's begin with the first of these reasons. Many of us strongly believe, particularly before being introduced to philosophy, that certain actions really ought not to be done no matter what anyone believes. Extreme examples best illustrate this. It seems obvious to many people that acts such as genocide and slavery are wrong, not because we do not like these acts or because most members of our community happen to dislike them, but because they are in themselves wrong. It will no doubt be noticed that this reason is not particularly sophisticated or impressive. I am basically just saying that realism has some inherent plausibility about it. But this is not necessarily an issue. Most of us

¹⁷ Some realists have sympathised with the idea that morality requires that moral truth be concealed from the majority of people. This possibility was most notably discussed by Sidgwick (1903, book 4, chapter 5) who claimed that it might be best, from his own utilitarian perspective, if utilitarianism's "truth" be known only to a small number of "enlightened" people who can be trusted departing from the rules of common-sense where it pays, in utilitarian terms, to do so.

are realists about an external world just because it seems highly intuitive (not that there aren't sophisticated arguments and against this view). Just because a view has some intuitive plausibility, it does not guarantee that it is correct. But it perhaps places the onus on those who object to the view to provide compelling reasons to reject it. Many anti-realist have attempted to do just that and I will consider some of these reasons in the course of this chapter.

The various tenets of realism each come with their own objections. Non-cognitivists tell us that although the surface grammar of our normative claims suggest they express beliefs, this is mistaken. There are no normative truths because normative statements are not truth-apt. Perhaps the non-cognitivist is correct. But if she is, it is worth noting that for many of us it does not feel that way from the inside. I know what it feels like to hold a belief. I also know what it feels like to express the non-cognitivist's non-doxastic mental states, whether that be to express an emotion, a commandment, or an endorsement of some set of norms. When I say or think that certain things are right or wrong or good or bad, the accompanying mental state I have feels to me to be more like a belief. This is what we might call the *phenomenological argument* against non-cognitivism. If normative judgments really do not express beliefs, it is at best strange that many people introspectively conclude that their normative judgments are belief-states. Unless we are under some kind of illusion, we are the epistemic authorities of the content of our own mental states.

Not only that, but normative judgments operate more like assertions of fact in that they can be incorporated into a chain of reasoning. A normative reasoner can argue that:

1. *If* failing to maximise happiness is wrong, *then* paying someone to not maximise happiness is also wrong
2. Failing to maximise happiness is wrong
3. Therefore, paying someone to not maximise happiness is wrong

The so-called *Frege-Geach Problem* for non-cognitivism is that it fails to explain how we can make sense of a chain of reasoning such as this.¹⁸ The cognitivist claims that normative judgments express beliefs, in which case, we can recognise the above argument as being deductively valid (even if, like anti-realist cognitivists, we deny that it is sound). This is hard

¹⁸ Also known as the 'embedding problem.' It was highlighted by Peter Geach (1965), though he attributes it to a point previously made by Frege. Non-cognitivists have proposed ways of avoiding this problem. See Gibbard (1990, p.95.)

to explain for the non-cognitivist because they need to give different semantic explanations for premises 1 and 2. Of premise 2, the non-cognitivist will claim that this statement is not truth-apt, but rather expresses a non-doxastic mental state. But this cannot be true of the first premise, because here the speaker is not expressing any kind of attitude, they are just claiming hypothetically that if we accept the antecedent, *that failure to maximise happiness is wrong*, then we must also accept the consequent, that *to pay someone to not maximise happiness is wrong*. Cognitivists suggest that this shows a failure on the part of the non-cognitivist to explain a key aspect of our normative discourse.

The second key motivation for realism comes from some of the deeply counterintuitive implications of anti-realism. Street is admirably honest in pointing out that her view - constructivism - implies that a fully coherent sadist - such as Caligula - whose strongest desire is to inflict suffering, may have decisive normative reason to do so (Street 2010). As I see it, this just seems too big-a-bullet to bite. It is too counterintuitive to accept a theory which tells us that provided the agents in question were fully instrumentally rational, Hitler and his fellow Nazis might have had decisive reason to bring about the holocaust or that Caligula might have had decisive reason to torture people.¹⁹ Similar counterintuitive implications arise out of relativism. According to the most basic relativistic view, one ought to act in accordance with the prevailing views of her society. But what if one lives in a society in which Nazism is the prevailing view? Does it thereby follow that I have sufficient reason to be a Nazi? Aside from that, relativism might even be self-defeating. For instance, what if the prevailing societal view is that we ought not to follow the prevailing societal view? Relativism seems to tell us both to follow and not to follow the prevailing societal view: a contradiction. Error theorists, like Richard Joyce, who believe that “all moral judgements are unjustified” (2013, p.351), can claim that Nazi moral beliefs are always unjustified. But by the same token, the error-theorist must also deny that anti-Nazi moral beliefs are justified.

Of course, these remarks do not go very far in addressing the worries of anti-realists. Non-cognitivists typically have very sophisticated reasons for believing that normative judgments are non-doxastic and subjectivism and relativism come in various forms, some of which less obviously have the implications I mention above. All I am doing at this point is outlining some of the most basic and fundamental motivations for accepting realism. Critics of realism, in my

¹⁹ Street is not claiming that Caligula’s beliefs are fully instrumentally rational. Just that if they were, then they would have normative reason to carry them out. She is also not claiming that Caligula has a *moral* reason to torture; just a *normative* one.

experience, typically grant that realism has some prima facie plausibility; their criticisms tend to be more philosophical ones pertaining to the realist's ontological, metaphysical and epistemological commitments, which are deemed by many philosophers to be unacceptable. The opposite, I think, is true of anti-realist theories, which despite not being as intuitively attractive as realism, come with more palatable metaphysical and epistemological baggage. This all means that, for the realist, the easy part is giving the positive case for their view; the hard part, is defending it against the various objections against it. Inevitably, as a thesis which focuses on one particular type of objection – the EDA – I simply cannot do justice to all of the criticisms that have been made. That said, behind EDA lie some of the very same concerns with realism that motivate a whole host of anti-realist views. I will consider some of these concerns at the end of the chapter, after having outlined the kind of realism that will be defended in this thesis.

2. Naturalistic Realism

EDA are generally considered to be problematical for all versions of realism. And whilst it is true that most EDA aim at realism generally, such arguments can be said to undermine different versions of realism in different ways. And according to some proponents of EDA, certain versions of realism are more obviously debunkable than others.²⁰ Some of the defences I will pursue in subsequent chapters are open to all realists. Others are not. This should not be surprising, because different versions of realism entail varying metaphysical and epistemic commitments, and insofar as EDA call into question such commitments, there is a need for the realist to spell out her view in detail before defending it. I will provide the details of my account of realism shortly. As a route into this discussion, I will now consider some alternative realist accounts which differ from mine and give my reasons for rejecting them.

There is a strong realist tradition of allying normative facts with natural facts, by claiming that normative facts are identical with, or in some sense reducible to natural facts. This view is known as *Naturalistic Normative Realism*.²¹ Naturalistic views are either *Analytic* or *Non-Analytic* (Synthetic). Analytic Naturalists²² claim that normative facts are true by definition,

²⁰ Street, for instance, claims that non-naturalistic realism is the most vulnerable form of realism to EDA.

²¹ Proponents of naturalistic realism include: David Brink (2001), Peter Railton (1986), Copp (2008).

²² For endorsements of analytic naturalism see: Jackson (2003) and Perry (1970). Perry is very explicit in tying his view to analytic naturalism, claiming that his view of value can be formulated using the equation: "x is valuable = interest is taken in x" (p.138).

meaning that when we say that some natural property n is some normative property n^* we are expressing a tautology not dissimilar to claims such as *bachelors are unmarried men* and *vegetarians are people who do not consume meat*. To illustrate, if the analytic naturalist is a hedonist, then she believes the claim *pleasure is the good* is true because the word *pleasure* has the same meaning as the word *good* or *the good* (Parfit 2011b, pp.266-267; Moore 1903, Ch.1). This view brings about a particular set of epistemological commitments. Like naturalists generally, analytic naturalists are *soft empiricists* about normative knowledge. They are empiricist in their belief that in being natural facts, normative facts are amenable to sense perception. They are “soft” empiricists because their account rests on the denial of the Quinean rejection of analyticity (Quine 1951). Moreover, in claiming that the relation between natural and normative facts is analytic, the analytic naturalist retains a role for a priori reasoning in our recognition of normative facts – albeit a role restricted to recognising synonymy between concepts.

In contrast, for the non-analytic naturalist, when I claim that n is n^* I am not making a claim about the meaning of words – that is, the natural property n is not said to have the same literal meaning as the normative property n^* . Rather, n is referring to n^* as a matter of contingent fact. In stressing this difference, non-analytic naturalists often draw upon the direct reference theory in the philosophy of language.²³ On this view, for our language to refer, it must successfully pick out or describe some aspect of the world. For instance, the statement that the longest reigning British Monarch was Elizabeth II is meaningful because the description of longest reigning Monarch correctly picks out Elizabeth II. But as noted by Kripke, these identity claims need not be analytic. It is not an analytic truth that Elizabeth II was the longest reigning British Monarch. Stock examples of such synthetic identity claims also apply to scientific discoveries, such as the discoveries that *water is H₂O* and *heat is molecular kinetic energy* (Brink 2001, p.160-162). The meaning of the concept *water* is roughly that clear, odourless, tasteless liquid which we drink and bathe in; only later was it discovered empirically that this stuff – in this world at least - happens to be comprised of H₂O. But clearly, H₂O was not built into our pre-scientific definition of water; and indeed, that water is H₂O seems only to be contingently true, for if in some other possible world a substance with all the causal properties of water happened to be comprised of something else, that substance would still satisfy the criteria for water. Non-analytic naturalists claim that the same is true when a normative fact refers to a natural fact. If hedonism is true, it is not because *pleasure* is

²³ Most notably: Kripke (1980) and Putnam (1962).

synonymous with *the good*, but because as a matter of contingent fact, the concept *the good* refers in this world to the natural property *pleasure*. This metaphysical claim also brings about a particular set of epistemological commitments. For if the identity relation between normative and natural facts is not analytic, then it will not be possible to ascertain which natural fact is picked out by our normative concepts a priori by merely by reflecting on what such concepts mean. To find out that *good* refers to *pleasure*, we must engage in empirical investigation.

Some philosophers claim that naturalistic versions of realism fare better than non-naturalist versions when responding to EDA.²⁴ This observation is in part motivated by a widespread scepticism about the supposedly “mysterious” metaphysical and epistemological commitments of the non-naturalist (which I will shortly discuss). It is also motivated by the thought that, if normative facts are natural facts, then we have a ready explanation for the relation between evolution and normative knowledge. Which is that evolution selected for organisms that are capable of forming true empirical beliefs because such knowledge helped our ancestors to survive. These empirical capabilities enabled us to form true beliefs about normative facts which, in being natural, are discoverable through sense perception. Whether a commitment to naturalism really does enable the realist to avoid EDA is questionable and several philosophers have rejected this proposed explanation (see Street, 2008). In any case, I will not be pursuing the naturalistic realist’s line of defence because I do not find it to be a convincing metaethical thesis, both in its analytic and non-analytic varieties. I shall now give some reasons for this belief.

2.1. The Case Against Analytic Naturalism

Analytic naturalism, insofar as it ever was a widely held view, largely fell out of favour following Moore’s criticism at the turn of the 20th century (Moore 1903, Ch.1). Moore claimed that analytic naturalists commit a “naturalistic fallacy” when they assert that normative and natural facts are analytically equivalent. He demonstrated this via the *Open Question Argument*. This argument begins with the following premise:

1. If some normative fact n^* has the same meaning as some natural fact n then anyone who fails to recognise that n^* is n is linguistically confused.

²⁴ E.g., Street (2006).

This is the claim that if two concepts are synonymous, their synonymy should be apparent to anyone who understands the meaning of the concepts. Consider the tautology that *bachelors are unmarried men*. Suppose someone asks whether I am a married man and I reply: “I am a bachelor.” Suppose the person then retorts: “I didn’t ask whether or not you were a bachelor, I asked if you were married.” For one who knows the definition of *bachelor*, their knowledge that I am one should be enough for them to infer that I am unmarried. Thus, in asking this question, this person is clearly linguistically confused. Put another way, the question of whether I am married - having already established that I am a bachelor - is *closed*, insofar as the answer is trivially obvious. If normative facts are tautological, it is reasonable to assume something similar would be true of them. That a basic apprehension of the words in question should enable us to recognise a priori that some n^* is n . If so, Moore’s first premise is reasonable.

Moore then denies the consequent of (1), by claiming that:

2. When people fail to recognise that some n^* is n they are seldom linguistically confused.

He thereby concludes that:

3. n^* does not have the same meaning as n .

The second premise is more controversial, but Moore gives some grounds for accepting it. Unlike the person who asks whether bachelors are married, people who inquire about whether some natural property is some normative property do not usually seem to be linguistically confused. For instance, if I were to inform someone that “I have had a *pleasurable* life” and they replied: “but have you had a *good* life?” it does not seem as though this person has failed to understand the meaning of either the word *pleasure* or *good*. Therefore, the question of whether *pleasure* is *good* seems *open*, in that it is not trivially obvious given the definition of the words. This does not mean that no natural property is good. Moore would accept that if *pleasure* is good, asking whether it is good may show ignorance of what is intrinsically valuable. But here the mistake is not a trivial one about what words mean, but a substantive one about what has ultimate value. The naturalistic fallacy is betrayed only by those who assert that natural and normative concepts have the same meaning.

Moore’s Open Question Argument counts against analytic versions of normative naturalism, even if it does not quite constitute a refutation of the theory. It could be argued, for instance, that Moore was wrong to assume that asking whether one half of a tautology is equivalent to

the other need constitute a closed question. This is the *No Interesting Analyses Objection* (Miller 2013, pp.15-16). Attempted analyses of contested concepts such as *knowledge* or *personal identity over time* may produce answers that are surprising to those familiar with the concepts (if Parfit's psychological criterion of personal identity turned out to be true, this would come as a surprise to many people²⁵). In response, I would suggest that many of these supposed analyses are not really about definitions, but about what our concepts refer to in a non-analytic sense. Even if it turned out that to be the same person over time depended on continuity and connectedness of psychological states, I am doubtful about whether this description could be said to be the *definition* of personal identity. Even if the *No Interesting Analyses Objection* is sound, and that strictly speaking, it is possible for an analytic truth to be non-obvious to competent language users, it still does not deal with the intuition that debates about normativity are about more than the definition of words. When characterising the awful moral beliefs of people like Hitler, surely we want to say that their moral failing is about more than their failure to recognise what words like *good* and *ought* and *wrong* mean. More plausibly, people who hold immoral views have roughly the same folk definitions of moral concepts as anyone else. They no doubt believe, as most of us do, that the concept *ought* refers to what people have sufficient reason to do under a certain set of circumstances, or that *the good* refers to whatever makes a persons like go well. Their mistake lies in making a non-linguistic, substantive error about what kind of things do, in fact make a person's life go well, and what kind of things people ought, and ought not, to do.

2.2. The Case Against Non-Analytic Naturalism

Moore thought that his open question argument undermined all versions of normative naturalism. This turned out not to be false. His argument only applies to naturalistic views which claim that normative facts are synonymous with natural facts; but as I mentioned, non-analytic naturalists deny this, since they claim that the identification of normative facts with natural facts is synthetic, akin to identity relations between concepts such as water and H₂O, and heat and molecular kinetic energy. The discovery that heat is molecular kinetic energy was informative because molecular kinetic energy was not contained in our folk understanding of heat. And if someone, despite knowing that there is heat in the room, failed to recognise that there is kinetic energy in the room, they would be making an empirical rather than linguistic error. Ditto, a hedonistic non-analytic naturalist will say, of people who sincerely ask whether

²⁵ Parfit's views on this can be found in part 3 of *Reasons and Persons* (1984).

a pleasurable life is the good life. Simply knowing the meaning of these words would not guarantee that someone recognised that the natural property of *pleasure* fixes the reference of the normative concept *the good* - this fact is known only by those with the necessary empirical experience; hence the openness of this question - or indeed any other question - which asks whether some natural property is good.

Non-analytic naturalism is vulnerable to other serious objections, however. For one thing, it is unclear whether the aforementioned descriptive identity claims are analogous to supposed identity relations between natural and normative concepts. Consider the example of water's being comprised of H₂O. Water is a physical substance. Even before we knew the further fact that it is comprised of H₂O, there was no ambiguity about what people were referring to when they spoke of water. To use the words of Brink (2001, p.160), one of the naturalists who cite such identity relations, water is the "colourless, odourless stuff found in lakes, rivers, etc. that is suitable for drinking, bathing, supporting life, etc." Now, I take the point that the pre-theoretic concept of water means something different from H₂O. My issue is that it is plainly obvious that our pre-theoretic concept *water* refers to something physical, given that we can see, touch and bathe in it – essentially, that it is amenable to sense experience – in a way that is not true of normative concepts. It is at very least controversial as to whether I can see, touch or bathe in *the good* or *the right*, or any other normative property (Gampel 1996, p.198). And indeed, the very fact that people so strongly disagree about what we mean by *the good* shows how it differs from things like water and heat, for although there was once serious scientific disagreement about what makes up heat and water, there was no real disagreement about what the words refer to in their non-technical sense (Gampel 1996, pp.195-196). This apparent dissimilarity raises serious epistemic problems when it comes to conducting the appropriate empirical investigation into *the good* and *ought* to establish what they are made of. In the case of water, we knew what physical 'stuff' needed to be investigated to ascertain its chemical composition. But since *the good* does not refer, in an obvious sense, to anything physical, how am I to know what I am using my empirical tools to investigate? And how am I to know what empirical techniques to use? (Need *the good* be inspected under a microscopic lens or be subjected to carbon dating?). So although I grant that synthetic identity relations of the kind described by the naturalist do exist, I strongly doubt that they can hold between normative and natural properties. Such a claim, though not a naturalistic fallacy, does seem to entail a category error (*the good* just is not that kind of thing).

2.3. Non-Naturalistic Realism

If naturalistic realism is false, as I have argued, the two remaining options open to the realist are *supernaturalism*, according to most versions of which normative truths are divine commandments issued by a deity or deities and *non-naturalism*, which holds that normative truths are neither supernatural nor natural. Supernaturalism rests on what many consider to be questionable metaphysical claims, such as the claim that Gods, or *God*, exists and intervenes in human existence by mandating certain acts as being right or wrong. But even if God did mandate some acts as being right and wrong, there is that separate question of whether this is because such acts are, *in themselves*, right or wrong – in which case, supernaturalism would not be incompatible with other versions of realism – or right and wrong solely in virtue of their being mandated by God – which would presumably make normative truths too arbitrary and changeable, for if God suddenly decided that altruism and kindness are wrong, then they would be wrong.²⁶ Option two, and the one to be defended in this thesis, is non-naturalism.²⁷ The non-naturalist thesis is part metaphysical, part epistemological. The metaphysical thesis pertains to what normative facts are. In answering this question, it is worth starting by saying what they are not. For starters, the non-naturalist claims that normative facts are not *natural facts*. The term *natural* is, of course, quite ambiguous and philosophers may mean different things by it. As I understand it, if some fact is not natural, then it has no spatio-temporal existence in the universe, unlike things like neurons, quarks, stones, and human beings. More precisely - though this is perhaps not implied by the last claim²⁸ - the non-naturalist's normative facts are not physical things. Now these claims alone are highly controversial, since the many philosophers who hold thoroughgoing naturalist or physicalist views will deny that any such facts satisfy these criteria. According to metaphysical naturalists, all facts are natural facts and according to metaphysical physicalists, all facts are physical facts. I will consider this objection to non-naturalism after I have further outlined the view.

²⁶ This line of argument is a variation of the *Euthyphro Problem*, attributable to Plato. In one of the Socratic dialogues, Socrates discusses whether piety is loved by the Gods because it is pious or whether it is Pious because it is loved by the Gods. See Plato (1914).

²⁷ See the following accounts of non-naturalism: Parfit (2011b, ch.31), Shafer-Landau (2003, ch.3), Moore (1903, ch.1) and Kaspar (2012).

²⁸ I suppose it depends what kinds of fact are exhausted by naturalism. There may be certain natural facts that are not physical facts. An example might be the subjective knowledge of *what it feels like to be conscious*, which, although contingent upon physical processes, is not itself, on some views, a physical fact. An example of such a view being Nagel's dual-aspect account of consciousness. See Nagel (1986, ch.3).

If normative facts are neither supernatural nor physical or natural, then what are they? Several non-naturalists have claimed that normative facts are *sui generis* or *in a class of their own*.²⁹ This would mean that in addition to the various categories of fact, such as natural, mathematical and logical, there is a distinct category of facts whose members include all and only normative facts – and within the genus of normative fact, there are several separate species including moral and prudential facts. The claim that normative facts are *sui generis* is plausible and I have no strong arguments against it, but I also believe that this understanding of them can be plausibly denied. Because I will have more to say on this in chapter 5, I will just baldly state my view. I think that normative facts are a species of philosophical fact, alike in both metaphysical and epistemological respects.³⁰ Metaphysically, when we think about candidate philosophical facts, such as Hume’s *Is/Ought distinction*³¹ (Hume 1985/1739, Bk.3, Pt.1, S.1), the claim that if there is no free will then there are no retributive grounds for punishment, Gettier’s ‘proof’ that the justified true belief account of knowledge is false (Gettier 1963), and the law of non-contradiction, such candidate facts seem to be highly abstract and do not, in any obvious way, denote physical properties.

Epistemic similarities also obtain between philosophical and normative facts. Philosophical facts are not, by and large, the kind of things that can be established through inductive reasoning. I do not know of a repeated physical event could serve to prove that no statement can be true and false at the same time. Normative facts, likewise, seem not to be amenable to this kind of justification, for even if every single person in history had tried to act in such a way as to try to maximise overall pleasure, it would not follow that this is what one ought to do.³² As with philosophical facts, the method through which we try to convince someone that some normative claim is true is through rational argumentation - using, in that respect, all the common-or-garden philosophical techniques, such as checking for internal consistency, considering the logical consequences of claims through thought experimentation, and ultimately, consulting our intuitions as the final court of appeal.

2.4. Normative Rationalism

The claim that normative facts, as a species of philosophical fact, are not primarily known through sense perception implies that normative knowledge is *a priori*. Those who accept this

²⁹ This is a view often associated with Moore. See Moore (1903, chs.1&2).

³⁰ For more on this understanding of normative facts, see Shafer-Landau (2006).

³¹ Hume noticed that no deductively valid argument can move from purely descriptive premises to a normative conclusion. I say more about this and its relevance to evolutionary ethics in chapter 2.

³² I will say more about this in chapter 5 in my discussion on intuitionism.

view about normative knowledge are rationalists and most philosophers endorsing a non-naturalist thesis about normative truths seem to fall in that category.³³ The hallmark of contemporary rationalism is the belief in the synthetic a priori (Bonjour 1998, ch.1) - and for normative rationalists in particular, normative facts are believed to be synthetic a priori truths. I have already discussed the idea, held by analytic naturalists, that normative facts are analytic and that knowledge of them is a priori to those who understand the concepts involved. Synthetic truths are not true by definition - which means we cannot recognise them purely based on our comprehension of the words involved. If normative facts are synthetic a prioris, then they are (1) known to be true through some kind of rational intuition and (2) not true by definition (that is, such truths are not analytic). Both of these claims are controversial. Many philosophers are sceptical about the idea that we can have intuitional or a priori knowledge.³⁴ Such an idea, they claim, is at best somewhat mysterious, at worst anti-scientific. However, some knowledge does seem best explained by rationalistic processes. For example, logical and mathematical truths cannot, in any obvious way, be established via induction. The only viable alternative to an inductive explanation is that we recognise such truths intuitively.

Even if there are truths that are not empirically known per se, but known to be true via some kind of intuition, it might be argued that all such truths are analytic, and therefore do not express substantive truths. In other words, it could be argued that all intuitive knowledge relates to propositions that are true by definition. This was the view of Ayer and other logical positivists (Ayer 1952, ch.4). Clearly, much of our a priori knowledge does consist of such analytic truths. My knowledge that if someone is a bachelor, then they must be unmarried, is simply a matter of my knowing that being unmarried is part of the definition of being a bachelor. However, there are some true propositions that can be known intuitively, but which do not seem to be analytic. I have already given examples of these in the form of mathematical and logical truths. Ayer argued that mathematical and logical truths are analytic rather than synthetic. (Ayer 1952, pp.74-77). However, it seems doubtful to me that in the case of $2+3=5$, $2+3$ has the same meaning as 5. The word 5 denotes a particular number whereas $2+3$ denotes the operation of adding the value of two numbers (2 and 3) together. Similarly, the concept *lowest prime number* does not mean the same thing as the number 2. 2 denotes a particular number. Whereas the concept *lowest prime number* denotes whatever number has the property of being the smallest of those which can be divided by 1 and itself – namely, the number 2. The number 2 has the

³³ For example, Parfit (2011b, pp.489-492), Sidgwick (1907, bk.4, ch.2), Ross (1930/2002).

³⁴ See Ayer (1952, ch.4).

property of being the lowest prime number, but the concept of *the lowest prime number* and *the number 2* are not analytically equivalent. In which case, the truth that *2 is the lowest prime number* is synthetic.

According to the non-naturalist, we should say something similar about normative truths. This would mean that if utilitarianism is true, then the proposition that *acting in such a way as to maximise expected happiness is right* is true, not because *acting to maximise expected happiness* has the same meaning as *right*, but because, whenever someone performs this action, their action has that separate, irreducibly normative property of being right. But here *being right* is a property of the action in question, it is not the same thing as the action in question – just as *being the lowest prime number* is a property of the number 2, but is not the same thing as 2. This is the nature of how the word “is” operates in normative propositions such as *x is good* or *x is right*, according to the non-naturalist. This is not a trivial point, because when we say that some *x is y* we often are making an analytic identity claim, such as when we say that *to be an unmarried man is to be a bachelor*. The non-naturalist uses the predicates ‘is good’ and ‘is right’ to assert properties rather than identities, analogous to statements like:

- (1) Everest is the world’s tallest mountain
- (2) 2 is the lowest prime number
- (3) The argumentative form *modus tollens* is valid
- (4) Rafael Nadal is the greatest male tennis player in history

Clearly, none of these claims is asserting an analytic identity relation. If, in discussing some philosophical argument I declare that “this modus tollens is valid!” I would not be saying anything remotely like “this modus tollens is modus tollens”; obviously so, since an argument can be valid without being modus tollens. I would be saying that an argument, which is a modus tollens, has the further property of being valid (in that its conclusion follows deductively from its premises). Now there is an important difference between the above claims. (1) and (4) are contingent claims. If, for instance, Novak Djokovic were to win several more tennis Grand Slams then he would be considered by most to be the greatest male tennis player. And if a meteorite obliterated Everest then K2 would be the world’s tallest mountain. (2) and (3) on the other hand are necessary in that their being true is not contingent on how the world is and therefore cannot be altered by future events. To make their account complete, the non-naturalist must take a stance on whether normative facts are necessary or contingent.

The matter of whether normative facts are necessary or contingent is a highly complicated one. It might be assumed that the question can be decided easily depending on whether one is a naturalist or non-naturalist. Since the naturalist claims that normative facts are features of the natural world, such facts are contingent on the nature of that world, and therefore need not hold in other possible worlds. The non-naturalist denies that normative facts are features of this world, thus presumably, their truth is not dependent on how our world is. However, this is too fast. Let's suppose that utilitarianism is a necessary truth. It might be argued that:

1. *If* utilitarianism is necessarily true, *then* prior to the big bang there was a moral requirement to maximise happiness
2. There was no moral requirement to maximise happiness prior to the big bang, because there was no one around to be bound by that requirement
3. Therefore, utilitarianism is not a necessary truth

Now let us assume that there were no extra-terrestrial beings capable of acting on normative requirements prior to the big bang. The question faced by the (utilitarian) realist is, if a requirement to maximise happiness has always been true, to whom did this requirement apply prior to the existence of sentient beings capable of acting on normative requirements? Now the realist might say that this requirement has always existed but only became *actionable* once rational sentient beings came into existence. This does, nonetheless, seem somewhat counterintuitive. Normativity is the domain of practical reason. For reasons to be binding there must be agents capable of recognising and acting upon them. The issue does not seem to be just that normative reasons are not actionable where there are no agents to respond to them, but that the very existence of reasons seems contingent on there being reasonable beings. If so, then it would not be the case that moral requirements are necessary in the sense that they are timelessly and unalterably true.

This thought might tempt us to say that if normative claims are true, then their truth must be contingent; at the very least, they seem to be contingent on the existence of rational beings. But it could be counterargued that even if normative truths cannot be necessary in the timeless and unalterable sense in which logical and mathematical truths are necessary, there might be a weaker sense in which such truths are necessary. The realist might argue that whilst it is true that the demands of practical reason are binding only wherever there are rational agents, wherever there are rational agents, the true normative principles are necessarily binding. The utilitarian realist might claim that wherever there exists the capacity to suffer, rational agents

have a reason, all things being equal, to alleviate, reduce and preferably eliminate that suffering. A Kantian realist may argue that wherever there are rational agents, it is necessarily true that such agents should never be used as a mere means. It could be objected that these claims sound contingent (e.g. on there being rational agents who can suffer). But this is not so. For although agents will have no reason to eliminate suffering in possible worlds in which suffering – or agents – don't exist, the utilitarian realist may still assert that her claims are necessary in the sense that there is no possible world in which it would be reasonable for rational beings to be indifferent about suffering where it does exist. And moreover, even if moral requirements, such as the utilitarian's, cannot be binding when there are no agents around to act in accordance with them, it still does not follow that there is no sense in which a true moral theory might be *timelessly true*. It might just be that the nature of that truth changes over time. For instance, the utilitarian realist might say that where rational agents capable of experiencing pleasure and pain exist, such agents will be bound by the moral requirement to maximise total pleasure minus pain. And although they might concede that no such requirement, as such, exists when there are no such agents, such as prior to the big bang, the counterfactual proposition that *if there were rational agents capable of experiencing pleasure and pain, then such agents would be bound by the moral requirement to maximise total pleasure minus pain* would nonetheless still be true.

Of course, this is just a rough sketch of how normative facts might be necessary. There remains a very strong objection to this claim related to evolution. It might be argued that normative facts cannot be necessary because their truth is contingent on the nature of the beings to whom they supposedly apply. To illustrate, if we are considering hedonism as a necessary truth, we might want to point out that pain's apparent badness and pleasure's apparent goodness, is contingent on the fact that human beings are animals which evolved to dislike and want to avoid pain and like and want to pursue pleasure (Street 2006, p.150). But if there were a species - a very alien species - which took a radically different evolutionary path from us, leading them to be totally indifferent regarding pleasure and pain - or even, let's say, to enjoy pain but hate pleasure - then it is hard to believe that hedonism could be true for them. This objection is a form of evolutionary debunking argument, and rather than discuss it now, I will wait until after I have said more about such arguments. In chapter 4, I will be considering the possibility that the judgment that pain is bad can be undermined by EDA. Because this feeds into the question of whether evolution undermines the necessity of a normative theory such as hedonism, I think that this the most fitting place to address this worry.

Now for a summary of the view to be defended in this thesis:

- 1) There are some normative facts
- 2) The truth of these facts is independent of the judgments of agents
- 3) All things being equal, humans have the ability to recognise these truths
- 4) Normative facts are neither natural nor supernatural. Like mathematical and logical facts, they are best described as non-natural. On one view, they are *sui generis*, meaning they occupy a class of their own. My preferred view is that they are a species of philosophical fact.
- 5) We come to recognise normative facts via rational intuition
- 6) In saying such facts can be recognised intuitively, I am not claiming that they are analytic. Normative facts are synthetic a priori
- 7) Normative propositions assert properties rather than identities. To say that *n* is good is not to say that *n* is identical with good, but that goodness is a property of *n*.
- 8) Normative facts are necessary; although their necessity differs from the straightforwardly timeless and unalterable sense in which, say, mathematical and logical facts are necessary.

2.5. Ethics Apart from Metaethics?

I have so far claimed that in responding to EDA, the normative realist must outline their metaphysical and epistemological commitments in relation to what normative facts are and how we acquire knowledge of them. Even this, however, is insufficient when responding to EDA for the following reason: whether one's account of realism is undermined by EDA depends in part on the realist's first order normative commitments. Many philosophers will find this claim highly questionable, owing to the widely held belief that metaethics and ethics occupy distinct areas of inquiry which can be pursued wholly independent of one another.³⁵ However, this widely held belief is mistaken and this becomes apparent when we think about whether EDA undermine realism. Because I address this fully in chapter three, I will keep this brief. The proponent of the EDA, or indeed, any kind of debunking argument, asserts that because a very large number of our normative beliefs are influenced by forces which do not track normative truths, we should be sceptical about whether such beliefs are true. Whether such a claim is successful depends on whether such distortive forces really did influence the

³⁵ A notable exception to this trend is Dworkin (2011, ch.2).

normative beliefs in question. Now let us imagine two different realists, both of whom hold different first order normative beliefs. Suppose that realist 1 believes that it is a moral imperative that one assigns greater moral weight to kin than to strangers. Suppose realist 2 believes that one should give equal moral weight to kin and strangers. Now let us suppose that we know, as a matter of fact, that the belief that we ought to give moral priority to kin is caused by some distortive force. Let's assume that no evidence exists that moral neutrality between kin and strangers was caused by any distortive process. Under such circumstances, it would follow that only realist 1's position is undermined by the relevant debunking argument and the reason has primarily to do with the nature of her first-order normative beliefs. Realist 2's position, on the other hand, is not undermined by a debunking argument, because the first-order normative belief she asserts to be true is not debunkable. Thus, whether the EDA will be successful against any particular realist will depend on whether her first order normative beliefs are debunkable.

What this means is that we might not be able to wholly defend realism against EDA without taking a stance on the content of first-order normative facts. Evolutionary debunkers, as I explain in the next chapter, make several different kinds of argument against the realist. Some debunkers choose to focus on the epistemic problem of explaining how we came to know normative facts; others, on how our normative capabilities and concepts evolved.³⁶ But some debunkers, notably Sharon Street (2006), emphasise how evolution shaped the content of our normative judgements. But if an argument like this undermines realism, it does so only indirectly though undermining the first-order normative beliefs posited by the realist. This requires, then, that in responding to EDA, the realist outlines her first order normative commitments. The position I hold is a broadly hedonistic utilitarian one. This is usually conceived of as the view that the right action is always to produce the highest total of pleasure minus pain.³⁷ I do not necessarily believe that utilitarian reasons and value claims are exhaustive of normativity, but it is the view which lurks in the background of this thesis and certain arguments - most notably in chapters three and four - lend support to this view. However, it is important to note that I do not think utilitarianism or any other form of consequentialism is the only normative view which is immune to EDA. And it goes without saying that, even if I am successful in showing that a combination of non-naturalistic realism

³⁶ Joyce (2006) and Kitcher (2011).

³⁷ Although I sometimes wonder whether our duty to minimise pain might be stronger than that to maximise pleasure. And for reasons which I outline in chapter 4, I take slight issue with the view that pain and pleasure are the only mental states hedonists should be concerned about.

plus a belief in utilitarianism avoids EDA, it does not thereby follow that utilitarianism is true. To show that hedonistic utilitarianism is the correct normative theory would require engaging in the more traditional territory of normative ethics, something which I have no intention of doing here. All I am saying is that EDA do not undermine the view of realists who accept a hedonistic utilitarian view.

3. Objections

This thesis is primarily concerned with one particular class of objections to realism and this means that there is not enough space to mount a thorough defence against the many charges that have been levelled against the view. But because many philosophers find some of the claims I have made to be highly questionable, I will briefly consider some of the primary objections. Some objections will be considered elsewhere in the thesis, where they have a bearing on the content of a particular chapter, such as *the argument from disagreement* in chapter 3 and the various arguments made against intuitionism in chapter 5.

3.1. The Argument from Physicalism/Naturalism

According to naturalists, all facts are natural facts. According to physicalists, of whom many are also naturalists, all facts are physical facts. It is possible for one to be a naturalist without being a physicalist. For instance, one might believe that, though it relies on physical processes, consciousness is not a physical fact, but rather, inherently subjective (Nagel 1986, chs.2&3). This would not count against the idea that consciousness is a thoroughly natural phenomenon. Both views deny that there is room in our ontology for non-natural facts. Therefore, normative non-naturalism is false. The most widely used defence against this charge - and the one which I am using - is the so-called *partners in crime* defence (see Kasper 2012, pp.65-66). This defence starts by denying both naturalism and physicalism on the grounds that there are counterexamples to such views. Stock counterexamples have already been mentioned, such as the truths of mathematics and logic. We can also throw in the facts of geometry, such as Euclid's axioms. As well as facts such as *nothing can be green and red all over*, whose categorisation is not straightforward. It goes without saying that it is a controversial question whether these facts are all indeed non-natural and a full defence of such a claim would require a thesis of its own. I deny that they are natural for several reasons. One is that such facts seem to be necessary, in that it seems intuitively true that there is no possible world in which both p and $\sim p$ are true at the same time, or where $2+2=[\text{something other than } 4]$. It is not just that it is

not the case in this world, but that it could never be case; and in not relying on the nature of our world, it is hard to see how such facts can be natural. Furthermore, it is difficult to see how such facts can be physical because we do not have contact with them via sense perception. I cannot ‘see’, in the sense relating purely to vision, that some argument is valid. Nor can I gain insight into some argument’s validity through touch, taste, smell, or hearing. The senses obviously play a role. I see the argument on the written page or hear it from another speaker. I may need to take a few lessons in logic to learn about the meaning of certain concepts. But in the end, my apprehension of some argument’s validity is a purely a priori matter.

The first part of the partners in crime defence is enough to deal with the argument above, for that argument intends to rule out the very possibility of non-naturalism about normativity on the grounds that no non-natural facts exist. But clearly, if that premise is false, then the argument is unsound. The second part of this defence makes the more positive claim that normative facts are themselves non-natural facts. This claim, of course, is not guaranteed by the mere existence of non-natural facts. It might be, for instance, that naturalism about normativity is true even if it is false about mathematics and logic. To substantiate this claim the realist needs to give an account of how normative facts are alike in respect to other non-natural facts. I try to give such an account in chapter 5 where I argue that normative knowledge derives from our ability to do philosophy. And therefore, I will save justifying the second aspect of the partners in crime defence until then.

3.2. The Charge of Platonism

The charge of Platonism bears certain hallmarks to the argument above in that it questions the metaphysical commitments of the non-naturalist; but does not rely on the truth of a thoroughgoing metaphysical naturalism or physicalism. According to this objection, the issue with non-naturalism is that it posits the existence of things that seem a bit mysterious.³⁸ Non-natural facts – so the objection goes – must occupy some odd Platonic realm that is separate from the natural world. And we humans that recognise these facts, must do so by some kind of mysterious intuitional insight into this realm. What precisely is this realm in which normative facts exist? And how is it that we humans came to have this peculiar ability to access it? These questions may lead one to think that the non-naturalist’s view comes close to supernaturalism. The issue is not so much that it denies hard naturalism or physicalism. But rather, that once we start talking about *existence*, it leads naturally to the question: exists *where*? Asserting the

³⁸ For examples of this charge see: Jackson (2003, pp.560-565).

existence of material objects easily answers this: such objects exist *here*, either in this world or this universe. But where do the non-naturalist's truths exist? If they are not in this realm, then in what realm are they?

This worry does have some force behind it and has been afforded substantial attention by non-naturalists. Derek Parfit, for example, in the second volume of his monumental defence of a non-naturalistic realist position, decided to drop the label of *realist* in response to the charge of Platonism. However, despite ceasing to call himself a realist, Parfit's position following this definition change would be classed as realist by most philosophers. 'Non-realist' Derek Parfit, for instance, maintained his belief in non-natural normative facts whose truth is independent of our judgments. Parfit's new preferred label for his view was *non-metaphysical cognitivism* (Parfit 2011b, pp.475-487). Interesting though it is, this semantic discussion is not what is important. I think Parfit's view is realist, he decided to reject the label; nothing much really hinges on this.³⁹ What is instructive, however, are Parfit's reasons for rejecting the realist label. According to Parfit, what distinguishes his view from other realist views, is that the normative facts he believes in do not come with any "positive ontological implications" (ibid, p.479). This phrase is, of course, slightly vague. For Parfit, whether some claim has such implications depends on the respect in which it is true. Sometimes we mean that the thing in question is real in the quite robust sense in which it is a physical part of the universe which causally interacts with other physical things; or that it exists in a 'different realm' such as heaven or some Platonic realm, with which we can have contact. But then there are truths which are not *real* in these senses. Again, we have the by now tedious examples of logical and mathematical truths. It also includes counterfactual claims. The statement that *if Hitler died in the Beer Hall Putsch, then Hitler would not have become chancellor in 1933* is of course true, but it is not a truth that can be said to 'exist' in the sense that it ever had a presence in the observable world (nor did it ever exist in some other 'realm'). The problem with the word realism, for Parfit, is essentially this idea that normative facts have a 'real' existence somewhere in the universe or outside of it and have causal properties. This would, in Parfit's view, be a metaphysically dubious claim. So for Parfit, we can do away with the claim that normative facts are *real* or *exist*, if by this it is meant

³⁹ The problem I have with the label "non-metaphysical cognitivism" is that it, on the face of it, seems indifferent about whether there are normative facts. Parfit believes there are such facts. But it seems to me that non-metaphysical cognitivism could also be used to describe John Mackie's error theory, which accepts the truth of cognitivism but denies the existence of normative facts on metaphysical and epistemological grounds. For a similar position to Parfit's, which does operate under the banner of realism, see Scanlon (2014).

occupy some place in space-time or some supernatural realm and instead just claim that some normative claims are true.

Parfit's response has, perhaps surprisingly, gone some way toward allaying the concerns of some naturalistically inclined metaethicists. The naturalistic realist Peter Railton, despite having once thought the idea of non-natural normative properties "smacked of Plato's heaven" (Railton 2017a, p.113) came to write that:

Reflecting on all this, I'm not sure how strenuously a Soft Naturalist should object, if at all, to "non-ontological" non-natural properties or to the non-natural facts attributing them. Soft Naturalists surely object to Platonistic Non-Naturalism, complete with an ontic conception of non-natural properties, knowledge of which is secured through a distinctive, "quasi-causal" faculty of synthetic a priori intuition, but Parfit's Non-Naturalism is not of this kind (Railton 2017b, p.54).

I think that Parfit's account should be broadly endorsed by the non-naturalist in responding to the charge of Platonism. I say broadly because I think, for reasons outlined by Larry Temkin (2017, pp.3-8), that Parfit's insistence that normative facts are causally inefficacious is too strong and arguably serves to undermine some of his other claims. What the realist should say, is that normative facts are causally efficacious in some ways but not others. The ontologically weighty properties that Parfit is distancing himself from are causally efficacious in the way that they interact with aspects of the physical world through physical contact, such as how the collision of two billiard balls may causally determine that one of them ends up in the pocket or how God may intervene in the world to punish the wicked. Of course, Parfit is right to claim that normative truths are not causally efficacious in this sense. But there is another, perhaps weaker sense in which they might be causally efficacious. Mathematical facts, for example, can be causally efficacious in the sense that they influence human behaviour. The negative reaction people have if they have been short-changed, for example, relies in part on their knowledge of mathematics. My declining to believe some argument might be causally related to my knowledge of validity. In scenarios such as these, it is not that there are mathematical or logical 'objects' which physically interact with us (there is no such things as a mathematical or logical *object*). But rather, that a person's knowing mathematical or logical truths can causally influence their behaviour. Parfit thinks that people can know normative truths. He also thinks, contrary to Humeans, that reason can compel us to act in accordance with these truths when it conflicts with our desires. However, if this is the case, then surely it follows that normative

facts are causally efficacious in the non-physical sense that, if a practically rational agent knows some normative truth, then this may be causally related to the agent's acting in accordance with that truth. For instance, Derek Parfit donated money to effective charities. He believed that people in well off societies have decisive normative reason to do likewise. Presumably, then, by Parfit's own lights, among the various causally relevant events leading to his decision to donate, they include knowledge of the fact that, all things being equal, well off people ought to make sacrifices in the interests of those less well off. Here as above, the claim is not that there are such things as moral 'objects' which can causally influence us. But rather, that normative and moral claims can be causally efficacious in the weaker, and non-Platonic sense in which possessing knowledge of such abstract truths may influence a person's behaviour.

Even if this response is successful, there is still this issue about how we gain insight into these non-natural normative truths. It is often claimed that for non-naturalism to be true, we would need some kind of "faculty", to borrow from Railton's passage above, in order to acquire normative knowledge. And according to critics, there is no cognitive faculty that allows us to detect non-natural truths. To quote the moral naturalist Frank Jackson: "our minds do not house detectors of the nonnatural properties Mooreans believe in [...]" (Jackson 2003, p.565). As I see it, there are two ways of responding to this charge. First, the non-naturalist might claim that the naturalist is just begging the question here. Without argument, naturalists often assert as though it were completely uncontroversial that humans do not have a cognitive faculty which enables them to detect non-natural facts. But this cannot be just stated as a truism; it must be argued for. Second, this argument sets up a straw target. Few non-naturalists claim that there is a specific "faculty" with which we get insight into normative facts (Kaspar 2012 p.32). G.E. Moore, the archetype non-naturalist to whom this belief is often attributed, was absolutely explicit in denying that he believed in a special normative faculty, as demonstrated in the following passage from his *Principia* (1903, p.x):

I wish it to be observed that, when I call [normative beliefs] "intuitions," I mean *merely* to assert that they are incapable of proof; I imply nothing whatever as to the manner or origin of our cognition of them. Still less do I imply (as most intuitionists have done) that any proposition whatever is true, *because* we cognise it in a particular way or by the exercise of any particular faculty.

Non-naturalists usually give a more sophisticated account according to which our recognition of normative facts is a by-product of well-established cognitive abilities, such as an ability to

reason – not that there is a specific normative or moral faculty. The repeated attribution of a belief in a special non-natural normative faculty on the part of the non-naturalist is an unreasonable one, given first that hardly any non-naturalists believe in such a faculty, and second, that we do not normally categorise thought in this way: few philosophers argue that we need a special philosophical faculty to do philosophy above and beyond our complex ability to reason.

Conclusion

Normative realism - and in particular, non-naturalism - largely fell out of favour from the early to mid 20th century, owing to the influence of, on the one hand, the linguistic shift, which put non-cognitivist accounts of normativity into the ascendancy, and on the other, the pervasive influence of thoroughgoing naturalist and physicalist views hostile to the idea of normative facts – particularly non-naturalistic normative facts. The linguistic shift has long since waned and although the hostility of naturalistically inclined philosophers remains, the signs are that non-naturalism is once again a metaethical theory to be taken seriously - a point underlined by the calibre of some of its contemporary exponents and the fact that some of its most prominent critics have softened in their opposition, to the point of expressing areas of potential agreement with the non-naturalist, the most notable of whom being the naturalistic realist Peter Railton and the expressivist Allan Gibbard.⁴⁰ In this chapter, I have tried to provide a comprehensive account of this view, one that will be the subject of my defence of realism against EDA in subsequent chapters. I have also provided some responses against some of the most pressing criticisms of non-naturalistic realism. I have argued that the view is not at odds with a scientific outlook, for it does not posit that there are normative objects that causally interact with physical properties, or that ‘exist’ in the sense of occupying some Platonic realm. My view asserts that there are normative truths, not normative objects; just as there are mathematical and logical truths but not mathematical and logical objects. The only sense in which such truths are causally efficacious is in the weak sense that knowing about them can influence human behaviour. In the chapters which follow, I will defend this view against the claim that realism can be debunked by EDA.

⁴⁰ See Railton’s (2017a) and (2017b) and Gibbard’s (2017).

Chapter 2: Evolutionary Debunking Arguments

Introduction

Since Darwin proposed his theory of evolution by natural selection, thinkers from a range of disciplines, from philosophy and evolutionary psychology to theology, have raised questions about its normative significance, particularly its relevance for morality. I will spend some time in the first part of this chapter outlining some of these historical developments, as some of the same considerations which troubled ethicists in the period following the publication of Darwin's theory bear on the contemporary debate over the metaethical implications of evolution. The primary focus of this chapter will be evolutionary debunking arguments themselves. Though the antecedents of EDA have a long history, their contemporary manifestation is quite unique. I want to provide an account of the nature of EDA and address questions relating to their claims, their structure, and their targets.

EDA are not specific to any single area of philosophy. They are discussed in both normative ethics, metaethics and epistemology.⁴¹ Although I am primarily concerned with the metaethical implications of EDA - especially insofar as those implications are supposed to undermine the view I wish to defend - I will have something to say about how EDA operate in normative ethics. This is for two reasons: the first being that an understanding of EDA in relation to normative ethics is imperative to understanding how they are used in a metaethical context. Secondly, in responding to metaethical EDA, it is necessary to engage at some level in normative ethics, largely due to considerations raised in the last chapter about how the two disciplines cannot be completely separated. Because my intention here is to characterise a particular class of arguments, most of what I have to say here will be unoriginal. The exception to this will be my characterisation of the distinction between the two most widely discussed EDA in the metaethical literature today; namely, the EDAs of Sharon Street (2006) and Richard Joyce (2006). I am not the only person to recognise that their arguments differ in key respects. But I do frame their differences in a way that has not, to my knowledge, been done before; and these differences are key, because they determine how the realist ought to respond to the argument. Having outlined the various EDA from which I will be defending the realist account

⁴¹ For a discussion of evolution and its implications for epistemology, see Street (2009b).

I outlined in the last chapter, I will then go on in chapters three and four to mount a defence against these arguments.

1. Evolution and Normativity

1.1.The Evolutionary Process

The theory of evolution by natural selection, first outlined by Darwin, is now accepted as scientific fact. Evolution explains how complex organisms adapt and change in response to their environment. In *On the Origin of Species* (1859), Darwin introduced the reader to his theory by discussing artificial selection, the process by which human beings engineer the reproduction of animals and plants to bring about desired characteristics (Ibid, ch.1). Should the desired characteristic be a dark species of canine, then the means of achieving this outcome would be to ensure that mating occurs between the darkest canines in the gene pool. Darwin recognised that the pressures which shape selection need not be artificial. In nature, some organisms are successful in reproducing; some are not. For traits to be passed on to future generations, the organisms bearing those traits must survive long enough to reproduce. In the short term, the influence of selection is usually small. The immediate descendent of the canines might, at most, be marginally darker than their parents. But a selection process of millions of years can allow us to make sense of vast variation, making comprehensible how species as variant as humans and cabbages share a common ancestor.

It is a mistake to think that evolution has a *goal* or an *end*. Such are the preserve of conscious beings in possession of intentions and motives. But the evolutionary process operates *as if* it had an end, and that end, is survival and reproduction. Darwin noted two primary means by which the selection process occurs: natural selection and sexual selection.⁴² Natural selection occurs in response to environmental pressures. Whilst on his voyage to the Galapagos Islands, Darwin observed that the various islands hosted distinct species of finch. He inferred that the finch shared a common ancestor, but that the varying local conditions - and therefore varying environmental pressures - to be found on each island resulted in divergence (Buss 2012, pp.4-5). This was because traits which enhanced survival in one island, were not fitness enhancing in others. And the traits which became most prominent in any particular island did so because

⁴² There are other means, such as *Genetic Drift* (Buss 2012, p.8). Darwin discusses natural selection and sexual selection in chapter four of *On the Origin of Species*. He gives the subject of sexual selection a more detailed treatment in chapter eight of *The Descent of Man*.

the organisms bearing those traits in the highest degree were most likely to survive and reproduce.

The other principal means of selection noted by Darwin is sexual selection. Darwin was famously horrified by the extravagant plumage of male peacocks.⁴³ He wondered why evolution would invest in something so apparently costly. It was not too long before he found a solution. The peacock's tail serves no obvious role in enhancing the organism's survival in response to its environment. Therefore, the adaptations resulting in ever larger tails cannot have occurred by means of natural selection. Rather, its utility derives from its role in soliciting the attention of mates (Darwin 2004/1879, pp.433-437). Female peacocks showed more interest in males with larger tails. Consequently, males with larger tails were most likely to reproduce and therefore, the size of the tail increased with each successive generation, eventually leading to its present size. Sexual selection occurs in two primary ways. The example mentioned above is *intersexual competition*, whereby selection occurs through the mating choices of organisms – in this case, the female peacock. It also occurs through *intrasexual competition*, whereby members of the same sex compete against one another to mate with members of the other sex (e.g., the combat which takes place between stags during mating season).

The evolutionary process is geared toward survival and reproduction, but of what? The consensus on this question has changed over time, but now has it that the gene - rather than the organism or the group - is the unit of selection.⁴⁴ This means that when we say that evolutionary pressures tend toward survival and reproduction, we are referring to genes. This insight was unavailable to Darwin, as he died before the advancements in genetics which shone light upon how traits are passed down through inheritance. Genes are tiny biological structures made up of DNA, which provide the information, or blueprint, which determines our physical traits (Stearns and Hoekstra 2005, pp.508-515). The physical expression of our genes is known as a phenotype (e.g., eye colour is primarily the phenotypical expression of the OCA2 gene). There are some key points to bear in mind. First, it is seldom the case that phenotypes are controlled by a single gene; rather, genes work in tandem to produce phenotypic expressions. Second, the assertion that the gene is the unit of selection does not imply a crude genetic determinism, whereby phenotypic expressions are *solely* the result of genes - environment also plays a prominent role (Buss 2012, p.17-18). And thirdly, just because biological traits are passed on

⁴³ In a letter to Asa Gray, dated the 3rd April 1860, Darwin wrote: "The sight of a Feather, whenever I gaze at it, makes me sick."

⁴⁴ Buss (2012, pp.12-17).

genetically, it does not follow that every facet of human behaviour is genetic - this point will become more prominent later when I consider whether normative thinking is one such facet.

There are several reasons why the gene-centric view has gained prominence. One is the scientific tendency toward reductionism which seeks to explain phenomena in terms of their smallest parts (Midgley 2014, Ch.1). But it is also because the gene-centric account provides a neater explanation of certain phenomena. Why is it that certain organisms are prepared to sacrifice their own evolutionary fitness for the sake of another? The social insects (bees, termites, ants) are notable for extreme acts of self-sacrifice, which do not make sense if the individual organism is the unit of selection; but make perfect sense from the gene-centric perspective because of the benefits they bestow on the altruist's genetic relatives (Dawkins 2016/1976, ch.10). In our species too, we see a strong motivation to protect one's offspring, even where that comes at a risk to the parent. Again, this makes sense if the evolutionary process is marked by genetic survival and reproduction.⁴⁵

1.2. The Relation Between Evolution and Normativity

The debate concerning the normative implications of evolution is almost as old as the theory itself. The publication of *On the Origin of Species* was met with outrage on the part of many Victorians by whom it was perceived as an affront on Christianity.⁴⁶ The issue for many religious people was their perception that the theory of evolution undermined the existence of God. And since God, according to most mainstream religions, is the source of morality, it follows that the truth of evolution would serve to undermine morality. The question of why the Darwinian revolution was unique among scientific revolutions in generating such extreme responses – unlike, say, the Copernican Revolution – are complex, but we can identify some key reasons. One is that in the Victorian era, a fundamentalist reading of the Bible was much more prevalent than in it is today and most Christians likely believed in the literal truth of the Genesis creation story (Richards 2000, p.20). The Young Earth Creationist's thesis that human beings were created at one fell swoop is plainly at odds with a Darwinian account which has humans as links on a chain stretching back millions of years. However, many religious people

⁴⁵ This is the idea of kin selection, see Hamilton (1964).

⁴⁶ Useful surveys of the initial response to Darwin's theory include Dennett (1995, ch.3) and Brooke (2009). As Brooke points out, the response on the part of Victorian Christians was not uniform. Some notable Christians, such as Charles Kingsley, were supportive of Darwin. And many were able to find reconciliation between Darwin's theory and creationism.

now reject a fundamentalist reading of religious texts and accept the broad contours of the theory of evolution, seeing it as part of the creation story, with God as the starter of the evolutionary process.

But there is another reason for religious hostility toward Darwinism – one less resolvable by appeal to metaphorical readings of the Bible. Namely, that unlike the Copernican or Newtonian scientific revolutions, which left the idea of *the soul* untouched, Darwin's theory clearly has implications for not only our bodily nature, but the nature of our mind (Dennett 1995, p.63).⁴⁷ The mere idea that evolution has any implications for human nature has proved a controversial one - both to religious and non-religious alike. Although I am sceptical of views which take this thought to the extreme - that is, by over-reducing human nature to evolution - the idea that evolution has played no role in shaping our nature is deeply implausible. The brain, like other organs, is a product of natural selection and therefore its functions were designed, along with all bodily functions, for the purposes of survival and reproduction; and insofar as mental phenomena is caused by the brain, we should expect the same evolutionary pressures to bear upon our mental activity (I will outline some evidence supporting this shortly, when I discuss some of the empirical support for evolutionary genealogies of normativity).

For now, I will leave open the question of the extent to which evolution has shaped human nature. The question I am now interested in is why this has relevance for normativity. According to one line of thought, one developed by Victorian contemporaries of Darwin, a Darwinian human nature is incompatible with being a moral person. One aspect of this worry is that if evolution is true, then human beings are not separate from the rest of the animal Kingdom, which might imply that we are destined to behave like animals (Brooke 2009, p.199).⁴⁸ Another worry relates to the struggle for survival inherent to the evolutionary process. That is, if we remove God from the picture of creation and replaced Him with evolution, we would be resigned to the fact that humans are selfish and incapable of acting morally. According to the reverend Samuel Wilberforce, a contemporary of Darwin's who participated in the 1860 Oxford Evolution Debate, evolution is not only at odds with science, but with something of "far more importance [...]." Namely, "the whole representation of that moral and spiritual condition of man [...]" (Wilberforce, 1860, pp.257-258). This bleak picture, it has to be said, has been encouraged by some philosophers and scientists sympathetic to Darwin, who

⁴⁷ Dennett (1995, pp.26-27) also argues that Darwin's theory undermined the *mind first* view of existence, which he attributes to John Locke (2004/1690, pp.550-551). According to Locke, it is "impossible to conceive that ever bare incogitative Matter should produce a thinking intelligent being [...]"

⁴⁸ Hence the caricatures of Darwin's theory which circulated at the time depicting humans as monkeys.

have overemphasised competition and struggle over other aspects of the evolutionary process, thereby painting a distorted picture of the traits which enhance an organism's survival.⁴⁹ More recently, the idea that evolution implies humans are selfish by nature owes itself, in part, to misunderstandings and misuses of some of the metaphors used to describe the evolutionary process. Although there is clearly a difference between selfishness and other anti-social traits at the genetic level, as opposed to the level of the organism, it is arguable that not everyone has borne that in mind in relation to *selfish gene* analogy.⁵⁰ Indeed, this distinction was arguably not born in mind by Dawkins himself, who bafflingly claimed in his book *The Selfish Gene* (2016/1976), that we should “try to teach generosity and altruism because we are born selfish” (pp.3-4) despite spending a chapter in the very same book – notably entitled “Nice Guys Finish First” – arguing that a gene-centric view of evolution supports altruism at the level of the organism!

There are two ways of interpreting this worry about evolution implying that human beings are selfish. One crude way of interpreting it is to read it as asserting that the idea that human beings are selfish and amoral alienates our sense of self. But of course, the Darwinian may simply respond: ‘Whoever said the truth should respect our sense of self?’ If evolution means that people cannot be moral, then so bad for *being moral*.’ A more plausible interpretation of the argument is that:

1. If an evolutionary explanation of human nature is true, then human beings are wholly self-interested and incapable of acting morally.
2. But we know that some humans are not wholly self-interested and do act morally.
3. Therefore, an evolutionary explanation of human nature is false

This argument has the quality of being valid. Not only that but barring extreme scepticism about the motives of altruists – e.g., impartial organ donors, those who die for a cause etc. – we have strong empirical reasons for accepting premise 2. Although many Darwinians – though not Darwin himself⁵¹ – led credence to the *nature red in tooth and claw* view of evolution,

⁴⁹ Even Darwin, who did recognise the importance of altruism and other prosocial behaviours in the evolutionary process, was liable on occasion to paint an overly Hobbesian picture of evolution, for which he received criticism from Kropotkin (1902, ch.1).

⁵⁰ Mary Midgley (2014, ch.1) has argued that Dawkins's selfish gene account supports a Hobbesian conception of morality grounded in self-interest; a view Midgley rejects, arguing that a properly understood Darwinism should lead us to reject social atomism. Now, as Midgley rightly points out, in the Dawkins quotation I include in the text, it does appear that he is using the term “selfish” to describe human motives.

⁵¹ As I shortly show, Darwin not only thought that evolution is compatible with morality but that it is the source of morality.

most contemporary evolutionary psychologists now believe that altruism and cooperation are part and parcel of a Darwinian account of human nature,⁵² a point which I will discuss at greater length shortly. If this is so, then the Darwinian could reject the first premise on the grounds that extreme selfishness would not be a consequence of an evolved human nature, and thereby deny the inference that moral behaviour is incompatible with an evolved human nature.

To clarify, the above worry is that if evolution explains human nature, and if an evolved human nature is a selfish one, then humans would be incapable of acting in accordance with the dictates of morality (assuming that the dictates of self-interest and morality are not coextensive). This worry goes away when we consider the empirical evidence that an evolved human nature would be one in which altruism is a key feature.⁵³ A separate worry relates to the implications which would follow from evolution being the *source* or *origin* of morality. It is a simple fact about human societies that human beings engage in normative activity, whether that be normative reasoning, such as thinking about what we ought to believe and what we ought to do; normative practice, such as issuing moral and normative commandments to others and setting up institutions to enforce normative codes. We are, as a matter of fact, a highly normative species and norms pervade just about every sphere of our lives. This raises the question about why this is so.

There is disagreement between some religious people about the connection between our normative – including moral – knowledge and God. On one widely accepted account, God is the source of morality, without Whose existence, we would have no moral reasons whatever. Furthermore, God provides human beings with the capacity to recognise these moral truths, in the form of scripture – such as in the case of The Ten Commandments – or divine revelation (or both). Why is it that a Darwinian account of human nature undermines this picture? First of all, if it is true that evolution played a role in shaping human nature, then we have reason to suspect its influence extended to our normative thinking. The first person to recognise this potential implication of evolution was none other than Darwin himself, who discussed the implications of evolutionary theory for human morality in *The Descent of Man* (2004/1879). In a chapter entitled “Moral Sense,” Darwin informs us that he shares the view, held by some of his contemporaries, that the most impressive feature which distinguishes humans from other animals is our “moral sense or conscience” (Ibid, p.120) However, although the moral sense is, in its developed form, a uniquely human trait, Darwin suggests that its origins are more

⁵² See Boehm (1999), Sober and Wilson (1998), Burkart et al (2018), Joyce (2006, ch.1&2).

⁵³ I will outline some of this evidence shortly.

ancient; and proposes an investigation into “how far the study of the lower animals throws light on one of the highest psychical faculties of man” (Darwin 1879, p.120). Darwin then goes on to make perhaps the earliest statement of an evolutionary genealogy of morality. Darwin writes:

The following proposition seems to me in a high degree probable – namely, that any animal whatever, endowed with well-marked social instincts, the filial and parental affections being here included, would inevitably acquire a moral sense or conscience, as soon as its intellectual powers had become as well, or nearly as well developed, as in man (Ibid, pp.120-121).

However, if this is so, then the ‘considerations’ underlying evolution’s shaping of our normative persuasions must relate to survival and reproduction at the genetic level. That is, evolution would have designed our normative nature in whatever way happened to maximise genetic reproduction. Clearly, this is at odds with the religious account raised above, according to which, moral commandments come from God: why would God tell us to do things just because they maximise genetic reproduction? (perhaps God’s commandments coincide with whatever normative beliefs are genetically advantageous, but this seems like a rather unlikely coincidence).

The early Christian critics of Darwinism felt that, on the one hand, the theory undermined the existence of a fundamentalist God; and on the other, that it provided an account of human nature at odds with religious morality. Rather than simply being a chapter in the history of Darwinian thought, we should see these early critics as being early proponents of a line of thought which dominates the debate about the connection between evolution and ethics to this day. Indeed, I will argue that their insights are crucial to the thinking behind EDA. One of these insights is ontological; the other, epistemological. Starting with the first, many philosophers nowadays basically accept the thought that, if we take God out of the picture, there is no room in our ontology for an objectively binding moral truth. A thought that was already widely held prior to Darwin, summed up in Locke’s claim that:

[...] it is hard to conceive, how there should be innate moral principles, without an innate idea of a *Deity*: without a notion of a law-maker, it is impossible to have a notion of a law, and an obligation to observe it (Locke 2004/1690, pp.93-94).

Similarly, Kate Manne suggests that appeals to objective moral truths do not have the force they might have had within a religious context. In discussing the force of moral concepts like “ought”, she writes: “Perhaps Anscombe was right that these terms have been emptied of much

of their sense, and have a ‘merely mesmeric’ force without a god in the picture” (Manne 2017, p.25). However, rather than see this as a reason to embrace God and reject Darwinism (Manne does not believe in God), many contemporary philosophers see this as a reason to reject realism in favour of an anti-realist conception of normativity, which, they argue, is a more plausible metaethical position in the absence of God. The epistemological point is this. If we accept a religious account of human nature, we may claim that God created us with the capabilities to recognise the dictates of morality. We could also say, if we believe in an interventionist God, that He or She guides us toward moral truth. But if we remove God from the picture, it raises the question of how we could know the difference between right and wrong. As Brooke puts it, this left Darwin’s contemporaries with the following question: “What was the ultimate ground of moral values if the evolution of the moral sense could be explained simply in terms of survival value, without reference to the transcendent?” (Brooke 2009, p.198). Although the anti-realists who endorse EDA would not frame the issue quite like this – they would not, for instance, buy into the idea that without God we cannot act morally – many would basically accept the thought that if evolution, rather than God, is responsible for our normative thinking, we would not be in an epistemic position to discover the kind of stance-independent normative truths that the realist believes in.⁵⁴

1.3.The Evolution of Norms: The Empirical Evidence

The claim that evolution is the source of our normative judgments is controversial. In addition to religious hostility to this claim, there is also scepticism among many philosophers and social scientists.⁵⁵ I will now give some considerations and provide some evidence to support the claim that evolution has had some influence on our normative judgments – though to reiterate, I do not believe that evolution is the only salient feature influencing our normative judgments. The initial considerations are quite straightforward. First, our bodily functions were designed by evolution for the purpose of genetic survival and reproduction. This consideration must apply to brains, and since mental activity is a function of our brain, it must apply to mental functions. Normative thought is a mental function and therefore, normative thought is likely influenced by evolution. Second, evolutionary considerations feature in any inference to the best explanation of certain aspects of conscious decision-making: it would be hard to fully

⁵⁴ Indeed, this is the thinking underlying Sharon Street’s (2006) EDA.

⁵⁵ For a discussion on hostility to adaptationist accounts of human nature within the social sciences, see Tooby and Cosmides (1992).

explain phenomena such as extreme kin altruism and sexual desire, without bringing evolution into the picture.

As a starting point, then, it is plausible that evolution has had some role in influencing human nature. However, it might be objected that it does not follow that evolution has had a hand in every feature of human nature, and therefore, we cannot be certain that it has played a role in shaping normative judgments. It is true that just because evolution has shaped some aspects of our mental life, it does not follow that it has shaped them all. But there are reasons for believing it has played some role in shaping normative judgments. The first thing to bear in mind is that normative behaviour is a human universal, by which I mean, it is not restricted to a particular culture. That is not to say each culture has the same normative beliefs and institutions - there are important differences - but it cannot be denied that all human societies are norm-based. The significance of this being that, if a trait is cultural, rather than part of our nature, then we would not necessarily expect it to be universal. But if a certain trait is hardwired into our nature, then we should expect to see it in all human societies (Tooby and Cosmides 1992, pp.78-79). Furthermore, there seems to be some normative judgments which are, to a greater or lesser extent, universally present in our species - I am thinking about, for instance, prohibitions on murder, incest, stealing; partiality toward close relatives and so on. Though of course, as a prelude to a later discussion, the fact that some normative judgments are not universal perhaps implies that the evolutionary account of norms is limited.

Belief in an evolved normative nature has been supported by work in fields such as evolutionary psychology, cognitive science, and primatology.⁵⁶ Starting with the latter, one of the ways of trying to establish if normativity evolved is to determine whether it has antecedents in our close evolutionary cousins. The thought being that if it has, then the origin of our normative judgments pre-date human culture and is to be found in our primordial past. There is strong evidence that many of the behaviours we associate with morality are found in primates. Perhaps the trait that is most associated with morality is altruism. I have already mentioned one form of evolutionary altruism in the form of kin selection. Many species of animal show strong attachment to their kin and are often prepared to bear costs to enhance their evolutionary fitness; behaviour which fits neatly with the human mantra that *blood is thicker than water*. In addition to the duty to be altruistic to kin, most think we have obligations to non-

⁵⁶ Jonathan Haidt's (2001) influential Social Intuitionist model of moral judgment draws upon evolution. I discuss Haidt's account more thoroughly in chapter 4. Joshua Greene has performed experiments showing that different parts of the brain correlate with certain types of moral judgments. Much of this work is summarised in his 2014 book *Moral Tribes*.

relatives. Although humans are the only animal with the notion of an obligation to their fellow species members, we are not the only species which exhibits altruism toward non-relatives. Grooming (checking for ticks) often occurs between non-related primates. The way this often occurs is that if one primate grooms another, the other primate will expect to be groomed in return (de Waal and Flack, pp.6-9). This is an example of *reciprocal altruism* (tit for tat). Other pro-social behaviours which have been documented include food sharing, third party intervention and conflict resolution (Ibid, pp.1-24).

In addition to behaviours and acts, other facets of normativity include emotions and concepts. These facets are usually intimately connected. The altruism that results in my giving to charity may have at its core a sympathetic attitude toward the plight of another. It may also have to do with my understanding of a moral concept like *fairness*. The precise role that emotion plays in morality is disputed. Some, such as *emotivists* and *sentimentalists*, place strong emphasis on the role of emotions in moral judgment.⁵⁷ Rationalists place more emphasis on the role of reason, though for some rationalists, the emphasis on reason is more in relation to how people ought to form normative judgment rather than how they do form them. Most philosophers grant that normative judgments and behaviour often do spring from emotions, most notably, empathy, sympathy; and in cases of policing norms, righteous indignation and anger. There is ample evidence that these *moral emotions* are shared by our close evolutionary cousins, most notably chimpanzees and bonobos, but also monkeys such as the brown capuchin.⁵⁸ Empathy and sympathy seem to be at play in recorded cases of primates consoling a fellow member of their group who has been attacked. In a famous study, de Waal and Brosnan (2003) found that Brown Capuchin monkeys exhibit something akin to righteous indignation when they are subjected to unequal distribution of food. In the study, the Capuchins who received the less desirable share of food witnessed another Capuchin receiving a more desirable share. Instead of accepting the lesser distribution, they threw it out of the cage in a fit of anger directed at the perceived ‘injustice’, suggesting they felt unfairly done by. In a more recent case, de Waal et al (2016) cite evidence that possibly suggests bonobos vocalise their disapproval when another member violates a “social expectation.” For example, bonobos are not usually surprised when they are subjected to aggressive attacks when they arise out of attacks initiated by the victims

⁵⁷ For an account of emotivism, see Ayer (1952, Ch.6). Sentimentalism differs from emotivism insofar as it is not necessarily a form of non-cognitivism (though sentimentalists can be non-cognitivists). Sentimentalists just claim that emotion is the driving force behind our moral utterances – it doesn’t necessarily take a stance on the semantic nature of those utterances.

⁵⁸ See de Waal (2000a); de Waal and Flack, 2000b), de Waal et al (2003, 2016), Burkhardt et al (2018).

themselves or out of competition for food, because aggression in these circumstances coheres with the norms of bonobo societies. However, de Waal et al found that bonobos are surprised when they are subjected to aggression of a kind which falls outside the norms of their societies – e.g., an unexpected or uninitiated attack – and vocalise their surprise and distress.

Before continuing, a caveat is in order. I am not asserting that primates possess anything remotely like a human moral sense. Even if capuchins exhibit anger when faced with an unequal food distribution, it is not necessarily underlined by a conceptual understanding of a concept like fairness. It is also up for question whether non-humans have normative beliefs. A human can believe in a propositionalised normative claim, whereas primates may well have something closer to a negative emotional response to it. Humans are also capable of moral argument. They use their highly developed rational capabilities to form new normative beliefs, as a logical inference from other normative beliefs (e.g. if you ought not to harm non-humans, then, since factory farming harms non-humans, you ought not to support factory farming.) We also (on our better days) strive for normative consistency. If I am a meat consumer who claims that we should not harm animals, I at least need to have a story to tell about how this can be reconciled with my support for the slaughter of the animals I consume. The basic normative sense possessed by our evolutionary cousins as well as our non-human ancestors is best conceptualised as a seedling which requires nourishment to fully flourish. It marks the origins of our human normative consciousness but cannot develop without the evolved rational and linguistic capabilities which are, perhaps more than any other attribute, those which go farthest in marking the divide between our species and others.

Evolutionary psychology is a discipline concerned with explaining psychological behaviour in terms of evolutionary adaptations. A key focus in recent decades has been on providing explanations for why normative practices evolved. Explaining an evolutionary function entails showing how that function enhanced the survival and reproduction of our ancestors from a gene centric perspective.⁵⁹ In the case of normativity, this can be done in several different ways. We might first ask the big picture question of why we evolved to be normative beings in the first place. Second, we might ask the more local question of why particular normative beliefs evolved. Let's start with the first question. It is a fact that human beings are a highly social species. For a range of cultural reasons, many present-day societies are significantly less sociable than hunter gatherer societies, which are smaller, more interdependent and more

⁵⁹ For a thorough overview of developments in evolutionary psychology see Buss (2012).

egalitarian (both in terms of the distribution of resources and power).⁶⁰ When trying to establish why normative thinking evolved, we should focus primarily on its utility in hunter gather societies, as these are the environmental conditions in which humans existed for 95% of our species' history (Burkart et al 2018, p.2). It is not only that being sociable was a feature of hunter gatherer life, but that it was a precondition for survival. It would be extremely difficult for one human to slaughter enough animals, or forage enough food, for their own and their family's consumption. Cooperation between members of the same band was essential, for it is far more efficient to hunt or to forage in groups and to distribute the proceeds than to be an individualist, but engaging in this kind of cooperation requires acceptance of and an ability to enforce norms.

There are several different reasons why norms play an essential role in cooperation and I do not intend to go into all of them here.⁶¹ But here are some of the main considerations. In order for cooperation to be effective, members need to do their fair share. A psychological propensity to retaliate against free riders is one evolutionary solution, but also, a set of agreed upon and enforceable norms including a prohibition on free riding (Buss 2012). Then there are questions about distribution. It would not be possible to engage in repeated cooperative endeavours if there was an unrealistic prospect that each participant will get a share of the spoils. These kinds of pressures may have led to the evolution of norms which determine the distribution of the benefits of cooperation, such as those relating to fairness and equality. A further consideration relates to game theory and in particular, the prisoner's dilemma. In any individual prisoner's dilemma, the most rational option, from a self-interested perspective, is defection; whereas from the level of the collective, it is best if everyone cooperates. However, the result changes when we consider iterated prisoner's dilemmas: a series of prisoner's dilemmas taking place over a length of time. Here, the rational choice from a self-interested perspective is cooperation, because defection would deter others from seeking your cooperation. From this we may infer that evolution selected for a tendency to engage in reciprocal altruism (tit for tat) as this would facilitate repeated instances of mutually beneficial cooperation.

Before moving on I want to make a clarification. It could be argued that because the underlying cause of some altruistic behaviours is a process geared toward the 'self-interest' of genes, such

⁶⁰ The remaining hunter gatherer societies provide us with some insight into how our ancient hunter gatherer ancestors might have lived; see Marlow (2005). For a thorough treatment of the egalitarian social arrangements of hunter gatherers and their causes see Boehm (1999, especially ch.2).

⁶¹ Burkart et al (2018) provides a more comprehensive overview of the forms of cooperation integral to hunter gatherer life.

behaviours are not *truly* altruistic. This is a mistake. Sometimes apparently moral acts are undermined by underlying selfish motives. If I nurse a parent at the end of their life to secure a favourable portion of their will, then although it appears altruistic, my action is selfish. However, we can also be altruistic without any conscious self-interested motive. When a parent cares for their children they generally do not do so based on a conscious calculation that such behaviour is likely to preserve their genes. Their psychological motivation is of love and concern for their children. Even if the ultimate reason we have these motivations is that they enhance reproductive fitness, this fact has no bearing on the conscious decision-making of organisms. Even if genes are *metaphorically* selfish, organisms need not *literally* be.

The big picture question of why we evolved to think normatively is largely answered by the fact that we are highly social creatures who rely on sophisticated forms of cooperation which can only be achieved by a system of norms. This answer also goes some way in addressing the local questions pertaining to the evolution of specific norms. We might ask how such norms aided us in our ability to cooperate. I have suggested how prohibitions on free riding and a basic sense of fairness and equality might serve such purposes. It has been argued that reciprocal altruism evolved to allow for cooperation, as we see in the case of chimpanzees, whose initial act of grooming is preconditioned by the expectation that it will be reciprocated. Outside of cooperation, we can also see how certain *normative principles* might have enhanced evolutionary fitness. Kin altruism increases the likelihood of one's genes passing into the future. Incest taboos make sense given that procreation through incest is more likely to lead to a recessive gene. Prohibitions on murder, stealing, unprovoked violence and other anti-social behaviours no doubt go some way in reducing the risk of premature death. It is debateable how far these explanations go. Some believe most, if not all, normative judgments can be explained in this way, such as E.O Wilson, who writes: "What, we are then compelled to ask, made the hypothalamus and limbic system? They evolved by natural selection. That simple biological statement must be pursued to explain ethics and ethical philosophers, if not epistemology and epistemologists, at all depths." (Wilson 1995, p.153). Others, such as myself, think there are limits to the evolutionary explanation of norms (as I will try to show in the forthcoming chapters).

2. Evolutionary Ethics

2.1. An Evolutionary Justification for Normativity?

Suppose evolution has played a strong role in shaping human normativity. We may wonder what follows normatively. One response might be to accept evolution's normative salience, but to take a hostile or pessimistic attitude toward this fact. An example of such a thinker was T.H. Huxley, who advised us to do all we can to fight against our evolved moral cognition. Huxley writes: "Let us understand, once and for all, that the ethical progress of society depends, not on imitating the cosmic process, still less in running away from it, but in combating it" (Huxley 1995/1893, p.134). Huxley's remark assumes – or at least, expresses hope that – humans are psychologically capable of correcting or reversing the corrosive influence of evolution on our normative beliefs. If that is so, then it would mean that the evolutionary explanation of normativity is not all expansive, at least not in the long term, should we choose to take heed of Huxley's recommendations.

Not all philosophers are troubled by the prospect of evolution playing a salient role in normative thinking. Some thinkers have found in evolution a justification for their own moral and political beliefs. Interestingly, this has historically resulted in thinkers with conflicting normative perspectives citing support from evolution. Herbert Spencer, the political philosopher who coined the phrase "survival of the fittest" (1884, p.444) saw in evolution a fierce fight for survival in which the strongest prevailed and the weakest were left to perish.⁶² Admonishing proponents of the Victorian poor laws, Spencer argued that to interfere with (what he saw as) the evolutionary process would stultify the progression of humankind, by advocating an "interference which not only stops the purifying process, but even increases the vitiation [which] absolutely encourages the multiplication of the reckless and incompetent [and] discourages the multiplication of the competent and provident [...]" (Spencer 1851, p.324). On the other end of the spectrum, the anarchist thinker Peter Kropotkin (1902) emphasised the role of altruism in evolution, in particular between members of the same species. Kropotkin argued that these natural altruistic tendencies bolstered the argument that a society based upon mutual aid and voluntary collectivism is possible.

⁶² At times Spencer conflates progress in a normative sense with evolutionary change. An example is his claim that humans are "the beings whose conduct is most evolved" (Spencer 1907, p.14). This conflation arises out of Spencer's apparent belief that evolutionary "progress" and moral progress coincide. But from a purely descriptive point of view, it would be a mistake to claim that humans are the "most evolved" species; we have just taken a different evolutionary path.

The idea that evolution can serve to justify moral or political viewpoints is highly controversial and for good reason. One issue is that such arguments are often not empirically supported. This seems true in the case of Spencer, whose political philosophy does not adequately account for the role of altruism in promoting evolutionary fitness. The other issue relates to how evolution is supposed to justify a normative claim. One may argue that a particular normative claim n^* is correct for reasons which have nothing to do with evolution, but claim that facts about evolution lend further support to our adoption of n^* as a principle. For instance, an anarchist might argue, quite independently from evolution, that political authority is illegitimate, and that the only just political arrangement is one in which there are no social hierarchies. Having established this as a worthy principle, the anarchist may then cite evolutionary considerations, such as an apparent innate altruism, in helping to prove that this is a *technically possible* political arrangement given the nature of human beings. Notably, in this case, evolution is not being used to justify the principle as such, but rather, the practical feasibility of adopting that principle. This is a wholly reasonable reason for the normative ethicist to draw upon evolution to support her normative claims. Indeed, it would arguably be irresponsible to advocate a moral or political system which humans are biologically incapable of following.

A more problematical line of thought entails arguing for n^* on the grounds that it corresponds, in some way, to what is natural in an evolutionary sense. This is the claim Spencer seems to be making in his remark that “The forgoing exposition shows that the conduct to which we apply the name good, is the relatively more evolved conduct; and that bad is the name we apply to conduct which is relatively less evolved” (Spencer 1902, p.19). According to this line of thought, it is not just that evolution shows that n^* can feasibly be adopted, but that it justifies n^* . This line of thought rests on the premise that if some action or trait is found in evolution, then we ought to emulate it. There are several issues with this line of thinking. One is that it seems to fall foul of Hume’s distinction between *is* and *ought* statements (Hume 1985/1739, Bk.3, Pt.1, S.1). One of the motivations behind citing evolution as a normative justification for some principle, seems to be that evolution is natural. But as Hume famously pointed out, the mere fact that something is the case does not in any way imply that it ought to be the case. It is an invalid argument to assert that because some act is natural that it ought to be performed, given the clear semantic difference between the descriptive “is” and the normative “ought.” Along similar lines, Huxley argued:

There is another fallacy which appears to me to pervade the so-called “ethics of evolution.” It is the notion that because, on the whole, animals and plants have

advanced in perfection of organisation by means of the struggle for existence and the consequent 'survival of the fittest'; therefore men in society, men as ethical beings, must look to the same process to help them towards perfection. (Huxley 1995/1893 p.132).

Therefore, even if human beings happen to be selfish or altruistic, it does not follow that we ought to be. However, this argument can be revised to comply with the is/ought distinction. One might argue, in the first instance, that the appropriate standard for ethical conduct is conduct which is natural for human beings. This might be the view, for instance, of *Deep Ecologists*, according to whom, nature's value is intrinsic and thus does not derive from its contribution to the welfare of human beings and other sentient animals.⁶³ If we accept that nature serves as a moral standard, we might argue (I am not saying that deep ecologists do make this argument), on evolutionary grounds, that:

1. One ought to act in accordance with nature
2. Maximising genetic reproduction accords with nature
3. Acting selfishly maximises genetic reproduction
4. Therefore, one ought to act selfishly

This argument, aside from whether it is credible, is at least valid. Unlike the argument that:

1. Maximising genetic reproduction is an act which conforms with nature
2. Acting selfishly maximises genetic reproduction
3. Therefore, one ought to act selfishly

The reason the first argument is valid is that the initial premise relating to acting naturally already contains an ought, meaning the conclusion does not move illegitimately from an *is* to an *ought*. However, it is also a questionable argument, because the first premise has seriously counterintuitive implications (and of course, because premise (3) is empirically false). Perhaps the strongest case against adopting nature as a criterion for right action was made by Mill (1874). Mill argued that, no matter how we define nature, it is not a suitable criterion for rightness. On the one hand, we might define nature as being all that has occurred naturally including human action. Clearly, if this is our definition, then any human action would be

⁶³ Of course, deep ecologists are also keen to stress that nature is not distinct from human beings; but that human beings have a place within, not outside nor above, nature. It should also be noted that adopting *evolution* as a criterion of rightness is not something usually associated with deep ecology. I cite this perspective just because it is one on which nature is claimed to possess intrinsic value. For a statement of deep ecology, see Naess (1973).

classified as natural; meaning one could not fail but to act naturally in whatever she does, rendering the criterion useless (Ibid, pp.15-16). Another definition would be to exclude human action from the definition of natural, so that, in acting rightly, we should imitate nature minus human behaviour. But the issue with this proposal is that although some features of nature cohere with our moral sensibilities, some do not (Ibid, pp.19-31). Altruism and pleasure are natural phenomena, but so too are diseases, natural disasters, and suffering. If nature is intrinsically good, rather than good because of its contribution to some other end, then it is hard to see on what grounds we can differentiate between these apparently desirable and undesirable features of nature when we try to imitate it for the purposes of right moral conduct. The lesson Mill draws is that acting in accordance with nature will be right only when it leads to desirable ends; in Mill's case, this will be when it maximises happiness; but one need not be a utilitarian to accept to basic point. A final issue seems to be this. Unlike in Spencer's time, we now know that the unit of selection is the gene. Thus, to say that some trait is desirable on evolutionary grounds is tantamount to claiming that it is desirable because it enhances genetic reproduction. But it is hard to see why rational people should adopt a normative principle which asserts this. If we did, we would hold a very peculiar consequentialist doctrine that would imply we should constantly be aiming to maximise our genetic offspring. I assume that a theory which has normative implications like these must be false.

Considerations such as these have led most philosophers to reject such a simplistic evolutionary ethical approach. There are, however, other ways in which evolution may serve to justify a normative principle. Some contemporary philosophers influenced by Aristotelianism have drawn upon evolution in their views on *the good life*.⁶⁴ According to teleological views like Aristotelianism, human beings are goal-directed animals whose wellbeing consists in exercising their human capabilities in a virtuous way (Aristotle 1926, Bk.1). The agent who accomplishes this can then be said to function in a specifically human way. This raises a question about what the capabilities whose development constitutes human functioning are. Aristotle was unable to give a Darwinian answer to these questions. One particularly influential Aristotelian account, which draws inspiration from Marx's early thought,⁶⁵ has been the "capabilities" approaches of Amartya Sen (1979) and Martha Nussbaum (2000). This approach claims that there are certain capabilities that, when actualised, enable us to function in a "truly human" way (Nussbaum 2000, p.72). Nussbaum is an objectivist about these capabilities.

⁶⁴ Notably Odenbaugh (2017).

⁶⁵ In particular Marx's earlier writings. E.g. Marx (1959/1844).

Although there may be cultural variation in the way we exercise these capabilities, the basic components of what it means to be human are cross cultural. Neither Nussbaum nor Sen offer an explicitly Darwinian account of the human capabilities, though some of those they mention clearly have a Darwinian origin. But their accounts do provide a blueprint for how an evolutionary approach to the good life might work. If it could be shown that the truly human capabilities have an evolutionary basis, and if the good life consists in exercising human capabilities, it follows that evolution determines what is good for us. (For a discussion of specifically evolutionary approaches to neo-Aristotelianism, see Odenbaugh (2017)).

Nussbaum provides one of the most detailed expositions of the capabilities approach. In her book *Women and Human Development* (2000), she claims that the good life for human beings consists in exercising the following capabilities (pp.78-80):

1. Life (living to the end of a normal human lifespan)
2. Bodily Health
3. Bodily Integrity
4. Senses, Imagination, and Thought
5. Emotions
6. Practical Reason
7. Affiliation
8. Other Species (“Being able to live with concern for and in relation to animals, plants, and the world of nature.”)
9. Play
10. Control over one’s environment (both material and political)

I have no issue with the claim that these capabilities are human - though many, if not all, can be found to some degree in other animals - nor with the claim that these traits contribute immensely to human welfare - though, as a utilitarian, I would claim that their value is instrumental. My issue is with the traits Nussbaum omits. One cannot have an account of human nature which leaves out the negative aspects of our nature. What about *jealousy*, *aggression*, *schadenfreude* and *selfishness*? One might retort that we should not include these traits precisely because they are undesirable. But the issue is, to differentiate between the desirable and undesirable traits, we need a criterion (e.g., the principle of utility). Nussbaum recognises that not all human capabilities are valuable, asserting that “Not all actual human abilities exert a moral claim, only the ones that have been evaluated as valuable from an ethical

viewpoint. (The capacity for cruelty, for example, does not figure on the list).” (Nussbaum 2000, p.83).

Later, in giving an account of how the capabilities achieve “political justification,” Nussbaum draws upon Rawls’s idea⁶⁶ of reflective equilibrium (Ibid, p.101). Briefly, if I am trying to establish whether *p* is a justified normative principle using reflective equilibrium, I test it against my background normative intuitions. If it coheres neatly with my intuitions, this is grounds for accepting *p*. If *p* conflicts with some of my intuitions, I need to decide whether to reject *p* or the intuitions with which it conflicts, until I reach a point of equilibrium between my intuitions and the principle I am considering. In relation to the capabilities approach, reflective equilibrium would presumably filter out the negative human traits because they fail to cohere with any decent person’s background normative intuitions. The first thing to note is that this response rests on reflective equilibrium being the best method of checking whether our normative beliefs are justified. I consider this question in chapter 5. But suppose reflective equilibrium is the best method. We might wonder why the good life consists in developing truly human capabilities. Why not just run any capability, whether truly human or not, through the test of reflective equilibrium? It seems to me that those endorsing a ‘truly human’ or natural standard of the good life face a dilemma. Either they should claim that the sole determinant of whether a capability contributes to wellbeing is whether that capability is truly human, in which case, they would need to bite the bullet and accept that apparently unattractive human traits are in some sense desirable. Or they might fully embrace reflective equilibrium, but in doing so, should not give automatic privilege to capabilities which are truly human (unless they think that only ‘truly human’ traits cohere with our background intuitions).

Another issue with this Aristotelian approach is that there are strong utilitarian arguments against it. First, as an explanatory thesis, utilitarianism does a better job of enabling us to differentiate between the negative and positive features of evolution. The utilitarian might argue that the desirable traits which Nussbaum mentions are likely to contribute to utility whereas the negative traits which I mention are likely to diminish it. The utilitarian can also more easily *weight* the various goods, namely, by assigning them value based on their contribution to utility. Third, the utilitarian could argue that the Aristotelian account does a poor job of accounting for the wellbeing of people who deviate from the norm. Suppose someone derives very little or no happiness from some of the capabilities Nussbaum mentions.

⁶⁶ See Rawls (1973, pp.48-55)

Would we still want to say that such a person enhances their wellbeing when she exercises them, or indeed, that in failing to exercise them, their wellbeing is the worse for it? Conversely, suppose someone derives immense joy from ‘unnatural activities.’ Need we say they are valueless because they are not “truly human”? I would suggest that a miserable life that makes the most of our “human” capabilities is less valuable than a blissful life which does not.

2.2. Evolutionary Debunking Arguments

The discussion so far suggests that attempts to ground normative justification in evolution have been broadly unsuccessful. The importance of this will be more apparent after I say more about EDA. As the title suggests, such arguments are said to debunk something; but what? Any belief in principle can be subjected to a debunking argument. In the case of EDA, the primary targets have been beliefs relating to normative ethics, metaethics, and to a lesser extent, epistemology. The purpose of a debunking argument is to show why the origin of some belief serves to undermine its credibility. For example, suppose that a friend has informed me that he will call me at 5pm. I look at the clock, which currently reads 4pm and I form the belief that my friend will call me in one hour. Now suppose that the clock is broken and that the hands are stuck at 4pm. This discovery should lead me to question my initial belief because its source is epistemically dubious. There are various ways this dubiousness might be cashed out. We might say that the fact the clock is broken means my belief was not *justified*; or that reading the time from a broken clock is not a *reliable* means of tracking the time.⁶⁷ Regardless of how we cash it out, it seems uncontroversial that my initial belief can be debunked by the knowledge that I read it from a broken clock. And yet, it might well be that my initial belief was correct.⁶⁸ For instance, it might be that the actual time just happened to coincide with the time stated by the clock. Or it might be that my friend called some time before 5pm which just happened to be an hour from when I misread the time. But this would involve an epistemic stroke of fortune. The fact that I have read the time from a broken clock does not guarantee that I have got the wrong

⁶⁷ According to reliabilists, a belief that *p* is justified only if *p* was formed by truth conducive (reliable) processes. See Goldman (2009, pp.336-335).

⁶⁸ I owe this example of the broken clock to Bertrand Russell (1964, pp.154-155). Russell provided the example of a person who just happens to tell the right time from a broken clock and claimed that it shows that having a justified true belief is not a guarantee of knowledge. The argument is also a precursor to Gettier’s (1963) famous rejection of the Justified True Belief account of knowledge. If I did not know the clock was broken, I arguably would have been justified in believing that the time shown by the clock was correct. Hence, I had a justified true belief that it was 4pm but I did not know it.

time; but it should undermine my initial *confidence* that I had the right time. The appropriate response to my discovery that the clock is broken is scepticism about the time it displayed.

These are the basic considerations behind a debunking argument, but they need to be formalised in order for it to operate as an argument. The basic issue seems to be that my belief was *caused* in an unreliable way, such that we have no good reason for expecting it to be true; and that if it turns out to be true, that is just a lucky coincidence. There are in fact several claims here, two empirical, one normative. The arguer might claim:⁶⁹

Causal Premise 1: You believe it is 4pm because you have read it from this clock, a clock which happens to be broken

Epistemic Premise 2: All things being equal, reading from a broken clock is not a reliable way to identify the correct time

Normative Premise 3: If a belief was caused in an unreliable way, then you should suspend judgment about whether it is true

Conclusion: You should suspend judgment about whether it is 4pm

This argument seems valid. The premises also appear to be true, except for the normative premise, which requires a caveat. Even if the original source of a belief is epistemically dubious, I might come to acquire further support for that belief from more reliable sources. Suppose that after looking at the broken clock, I looked at my phone, which also happened to show the time as being 4pm. In which case, I could respond to the debunking argument by saying that although I accept that the initial cause of my belief is dubious, I can appeal to other reasons for why that belief is justified; namely, that the clock on my phone is working and that also read 4pm at the time in question. We could then revise the premise as follows:

Revised Normative Premise: If a belief was caused in an unreliable way, then unless there are independent grounds for that belief, we should suspend judgment about whether it is true.

Debunking arguments can come in local and global forms. A local debunking argument seeks to undermine confidence in a particular belief, such as the belief that it is 4pm. A global debunking argument seeks to undermine a whole category of beliefs. Suppose that I am indoctrinated into some strange cult, in which I am informed that the planet we live on is called Berth and that politicians are lizards; a small number of whom are in control of everything.

⁶⁹ I take the first two premises, which I have modified slightly, from Kahane (2011, p.106).

Suppose I discover conclusive evidence that the person from whom I acquired this world view is a known charlatan. This fact should not only cause me to question some particular Berth-related belief, but all my Berth-related beliefs. We might devise the following debunking argument:

Causal Premise 1: You hold all your Berth-related beliefs because you were told them by Bert, who is a known charlatan

Epistemic Premise 2: All things being equal, accepting the claims of a charlatan is not a reliable guide to truth

Normative Premise 3: If a belief was caused in an unreliable way, then, unless there are independent grounds for that belief, you should suspend judgment about whether it is true

Epistemic Premise 4: There are no independent grounds for believing any of your Berth-related beliefs.

Conclusion: You should suspend judgment about your Berth-related beliefs

The EDA⁷⁰ is a specific kind of debunking argument, one which, in drawing attention to dubious origins, refers to evolution's role in the formation of a belief. At its core, it is the claim that in the case of some beliefs, if they are caused by evolution, then we have reason to doubt their truth. I say "some beliefs" deliberately, because not all beliefs are undermined by having evolution as their cause. This is a point worth clarifying. Something may be a questionable epistemic source in the case of some beliefs but not others. If Don lies to everyone except his mother – to whom he is always truthful – then beliefs derived from Don's testimony are questionable only when they are not held by Don's mother. Likewise, beliefs derived from evolution might in some cases be questionable, in other cases less so. Of course, we need an account of why this is. In the case of evolution, some argue that certain kinds of knowledge capabilities evolved because they enhanced our ancestors' survival and reproduction (Joyce 2006, pp.182-183). Basic perceptual knowledge helped us navigate our surroundings. Deductive and inductive reasoning led us to form various advantageous true beliefs, e.g., the fact that all previously observed tigers have tried to maul me gives me reason to think the next one will. The same is also probably true of mathematics (ibid). If, as I have already suggested, norms relating to equitable distribution of resources evolved, knowledge of mathematics would be required to enable us to decide what is equitable. Such knowledge also aided with avoiding

⁷⁰ For a wide-ranging discussion of EDA see Kahane (2011).

risk. If I saw four tigers enter a forest and later saw three leave, I can deduce that there is at least one tiger left in the forest.

The reason why an evolutionary explanation for such knowledge does not serve to undermine them is that their truth is integral to why they were selected.⁷¹ What I mean is that evolution would not have an interest in us believing any old mathematical or perceptual proposition: see how far an organism would get if she thought tigers were herbivores, that Hemlock is a nutritious beverage, or that the number of tigers in the forest is zero and not one. Of course, this explanation has its limits. It might well be that knowing basic arithmetic aided our ancestors' survival, but it is less clear how knowing Pi would have. Similarly, it is plain to see how being able to predict the likelihood of an organism attacking me would be evolutionarily advantageous, but it is less clear how knowing about quantum mechanics would be. However, this is not necessarily an issue for an adaptationist explanation of mathematical and perceptual knowledge. Evolution provided us with the ability to recognise certain basic mathematical and perceptual truths, but our advances in these areas involve the *extension* and *application* of these abilities in ways they were not necessarily 'intended for'.

It follows that some of our beliefs are not undermined by their evolutionary origins. How, then, can evolution serve to undermine some beliefs? In the above examples, it would be advantageous for our species to have true beliefs about the branch of knowledge in question. As such, we might suppose that evolution's influence on such beliefs would tend in the direction of truth. However, this is possibly not the case when it comes to some other categories of belief, such as our normative beliefs. The basic thought is that it is hard to conceive of how

⁷¹ This is not without question. For one thing, there is widespread dispute within the philosophy of mathematics about whether mathematical claims are true in the realist's sense of being mind-independent (see Field 2016). Moreover, some philosophers take issue with the argument that EDA do not affect mathematics in the way they do ethics (this is relevant, for if true, it would presumably undermine the partners in crime defence outlined in the previous chapter, for mathematics would no longer be a branch of non-naturalistic knowledge). They claim that to explain why we hold the mathematical beliefs we do, we need not presuppose that they are true. Clarke-Doane (2012), for example, holds that in the kind of scenarios I describe above showing why mathematical truths would be advantageous, it is actually more to do with the advantages of knowing true first-order logical truths than the truths of mathematics (as seems to be the case in the example I gave in which someone deduces that the number of tigers in the field is 1 and not 0). Even if this were true, which I doubt, the realist could concede anti-realism about mathematics, but solely focus on the analogy between knowledge of logical and normative truths, thereby keeping the partners in crime defence alive, albeit with one less partner (Clarke-Doane's argument that mathematical realism is immune to EDA seems to presuppose realism about logic). But the evolutionary benefits of knowing mathematical truths can be boiled down to the benefits of knowing logical truths. Take the example of distributing food. In order for our ancestors to cooperate in hunts they would presumably need to anticipate they will receive a fair share of the spoils (e.g. if there are 12 people participating in the hunt they will need to be able to divide the spoils between 12). In cases like this, the mathematical calculations would need to be correct in order to get the intended benefit. Thus, in order to explain how our ancestors benefitted from believing in certain mathematical propositions, we need to presuppose the truth of those propositions.

an ability to recognise normative truths could enhance evolutionary fitness. Of the evolutionary redundancy of normative truths, Street writes: “a creature obviously can’t run into such truths or fall over them or be eaten by them. In what way then would it have promoted the reproductive success of our ancestors to grasp them?” (Street 2006, pp.130-131). Of course, one cannot fall over or be eaten by mathematical or logical truths; and thus, the idea that normative truths are not causally efficacious in a physical sense – if this is the point Street is making – is not in itself an issue. But there remains a problem. We can give plausible reasons for why mathematical knowledge would be advantageous from an evolutionary perspective. It is hard to give a similarly convincing account in the case of normativity. If we were to claim that normative knowledge evolved, we would need to claim that our ancestors who *knew what they ought to do* reproduced at a greater rate than those who did not.

One option here is to bite the bullet. We might argue that normative knowledge did enhance the evolutionary fitness of our ancestors. This might be the view of those philosophers who ground their moral and normative theories in evolution. But I have given several reasons for why such a view should be rejected. Another reason for doubting that normative truths are the kind of things which would be evolutionarily advantageous is that several plausible prospective normative truths do not seem to enhance evolutionary fitness in any way. For instance, suppose that a trolley is hurtling toward five people on a track.⁷² I am standing on a footbridge situated in between the trolley and the five people, under which the trolley will pass. Rather than allow the five people to be killed, I jump into the path of the trolley, sacrificing my own life for the five. I think most people would characterise this action as morally admirable, even if few other than utilitarians would consider it obligatory. What is beyond question is that unless I am genetically related to some of the people on the track, such behaviour is not in my evolutionary interest – indeed, it is *against* it. This is an extreme example. A more realistic example is whether I ought to reduce my carbon footprint. It is highly plausible that I have a reason to do this. Of course, in choosing to do so, it is unlikely that I will personally benefit, since the positive consequences of such actions will not come about for many years. One might try to rationalise this action in terms of the consequences for future people to whom I am genetically related. This is fair enough to a point. But we surely wouldn’t want to say that the moral obligation to reduce carbon emissions is contingent on whether it will affect one’s genetic relatives and that, for instance, life-long bachelors whose families have all died have no such obligation. This all suggests that knowing and acting upon plausible moral beliefs will in many

⁷² The Trolley Problem was first discussed in a paper by Philippa Foot (1967).

cases not enhance one's evolutionary fitness – indeed, it will often be against it – and therefore, if we suppose that some of these plausible moral beliefs are facts – as realists do – then we should deny that it is in our evolutionary interest to know such facts.

The suggestion, then, is that evolution would not select for an ability to acquire normative knowledge. However, many philosophers and evolutionary psychologists argue that evolution selected for the capability to think in normative terms and some claim that evolution selected for certain normative beliefs. If evolution did not select for us to hold certain normative beliefs because they are true, then why were they selected? According to Sharon Street's *adaptive link* account of normativity, evolution selected for certain normative beliefs because they enhanced our ancestors' survival and reproduction (Street 2006, pp.126-129). The claim is not that holding *true* normative beliefs would enhance reproductive fitness. The point is that some normative beliefs – irrespective of whether they are true – aided our ancestors' reproductive fitness. In asking why we hold such beliefs, such as the belief that we should be kind to kin, we need not suppose, as we do in the case of science and mathematics, that the reason has something to do with the evolutionary value of knowing truths in these areas. We need only suppose that the members of our species who held beliefs of this kind were more likely to pass on their genes. The key point here is that if evolution selected for a normative belief n^* , it did so for reasons completely unrelated to its truth. Like the broken clock for time telling, evolution is not a reliable truth-tracking mechanism when it comes to normative beliefs. Insofar as it has guided us normatively, it is most likely not in the direction of truth.

Now that I have outlined the basic thinking behind EDA, I am going to put it into argumentative form. Debunking arguments come in local and global varieties. A local EDA seeks to undermine a particular normative belief on the grounds that such a belief was caused by evolutionary forces and these forces are not a reliable guide to normative truth:

1. You believe you should prioritise the interests of your kin over strangers because members of our species evolved to hold this belief
2. Since normative knowledge is not fitness enhancing, evolution is not a reliable guide to forming true normative beliefs
3. If a belief was caused in an unreliable way, then unless there are independent grounds for accepting that belief, we should suspend judgment about whether it is true.
4. There are no independent grounds for believing we should prioritise the interests of kin over strangers

5. Therefore, we should suspend judgment about whether we should prioritise the interests of kin over strangers

The plausibility of all these premises - besides premise 4 - have already been defended. It might well be that there are independent grounds for holding the belief that we should prioritise kin. And of course, this argument does not count against the idea that kin partiality is indirectly morally required for its contribution to an intrinsically correct normative theory: say, if we could independently verify utilitarianism and then show that kin partiality generally maximises happiness. Note that the target of this argument is one who accepts the first-order normative claim that kin partiality is morally required. The argument does not appear to have any metaethical implications. A normative realist may accept this argument, especially if she does not think that kin partiality is one of the true moral requirements. Thus, it is not surprising that normative realists often use local EDA. One example is Peter Singer and Katarzyna de Lazari-Radek (2014, ch.7), who developed an EDA against ethical egoism. Another is Derek Parfit, who has argued that our retributivist tendencies can be explained by evolution and that such an explanation undermines their justification (Parfit 2011, p.429).

Kahane goes so far as to claim that the use of EDAs presuppose normative objectivism. From a realist perspective, the point of normative discourse is to state objective facts, “But if there is no attitude-independent truth for our attitudes to track, how could it make sense to worry whether these attitudes have their distal origins in a truth-tracking process” (Kahane 2011, p.112). However, this claim is too strong. An anti-realist may deny that there are objective normative truths whilst believing that we ought to make normative judgments in a rational way. In which case, they might see evolutionary forces as undermining our rational capabilities. Similarly, “best opinion” versions of anti-realism will presumably downgrade opinions that are grounded solely in our evolutionary instincts.

EDA have metaethical implications when issued globally. Perhaps the most influential argument of this kind was made by Sharon Street. According to Street, “the forces of natural selection have had a tremendous influence on the content of human evaluative judgements” such that “our system of evaluative judgements is thoroughly saturated with evolutionary influence” (2006, pp.113-114). Street does not put a precise figure on the proportion of our normative beliefs that were shaped by evolution. Nowhere does she state that every single normative belief evolved. And she does concede that factors such as culture and rationality have some influence. But in claiming that our normative judgments are “thoroughly saturated”

by evolution's influence, we can presume she thinks that the overwhelming majority of our normative beliefs were influenced by evolution. Based on this premise, we can reconstruct the first premise of the local EDA to make it global. We would then have the following argument:

1. You hold most of your normative beliefs because members of our species evolved to hold such beliefs
2. Since normative knowledge is not fitness enhancing, evolution is not a reliable guide to forming true normative beliefs
3. If a belief was caused in an unreliable way, then unless there are other grounds for accepting that belief, we should suspend judgment about whether it is true.
4. There are no independent grounds for holding most of our normative beliefs
5. Therefore, we should suspend judgment about whether most of our normative beliefs are true

The upshot of this argument is not that we should be sceptical about any particular normative belief being true, but about most of them being true. The normative realist claims that there are normative truths and that we can know them. But if most of our normative beliefs come from an unreliable source - at least, unreliable if the aim is to know stance-independent truths - then arguably realism as a metaethical position is undermined.

This is not the only way of formulating the global EDA. Indeed, it is not quite the way Street formulates it - for one thing, Street does not put the argument in formal premises and conclusion. Street proceeds by arguing for the first premise, the one which states that normative beliefs were mostly shaped by evolution. After which, she claims that the realist must take a stance on the type of relation evolution bears to these truths (Street 2006, p.121). This is Street's *Darwinian Dilemma* and the realist's two options are either to assert or deny a relation between evolution and normative truth. To assert a relation would be to claim that knowing normative truths would be fitness enhancing and claim that this explains how our normative beliefs came to converge around truth. Taking this option faces the scientific objection that normative truths - if they exist - are not the kind of things that evolution would invest in, hence the reason Street prefers the adaptive link account. The other option - to deny a relation - involves the realist accepting that evolution did not select for an ability to recognise normative truths, whilst maintaining that somehow we ended up discovering them anyway. This response faces the *massive coincidence objection* (Ibid). Given that here the realist would essentially be accepting that most of our normative beliefs were shaped by a process that does not track normative

truths, only an incredible stroke of fortune could have ensured that we came to hold true normative beliefs. Street thinks such a coincidence is hard to believe, likening it to trying to find Bermuda with nothing but the help of the wind and tides. The upshot being that however the realist responds to the Darwinian dilemma, their account faces unsurmountable difficulties. And therefore, we should reject realism and opt for an anti-realist alternative.

2.3.The Distinction between Content Partial and Content Neutral EDA

Around the same time that Street published her argument, a similar line of argument was developed by Richard Joyce in his book *The Evolution of Morality* (2006). However, their arguments differ in some key respects. Street is explicit in claiming that evolution has shaped the content of our first order normative beliefs. Indeed, she draws up a list of some of the beliefs which she believes are prime candidates for an evolutionary explanation, among which she includes things like kin partiality and the belief that pain is bad. However, not all philosophers who endorse EDA accept the claim that evolution shaped the content of our normative beliefs. Some argue that although evolution gave us the tools necessary to think in normative terms, or the *normative capabilities*, it did not necessarily shape the content of the first-order normative beliefs we came to hold. I call this the distinction between content neutral and content partial evolutionary genealogies:

Content Partial Evolutionary Genealogy: (1) *the human ability to engage in ethics has an evolutionary basis.* (2) *Evolution shaped the content of our first-order moral beliefs.*

Content Neutral Evolutionary Genealogy: (1) *The human ability to engage in ethics has an evolutionary basis.* (2) *Evolution did not shape the content of our first order moral beliefs.*

Joyce falls within the second of these two categories, as demonstrated by his remark that: “[the] claim was just that there is a specialised mechanism [...]” which “comes prepared to characterise the world in morally normative terms; moral concepts may be innate even if moral beliefs are not” (2006, pp.180-181). Joyce argues that morality is the culmination of various traits that enhanced our ancestors’ evolutionary fitness. He spends a chapter arguing for the biological basis of altruism (Ibid, Ch.1), citing considerations discussed earlier in this chapter.

He also thinks the belief in categorical reasons or imperatives was an essential component in the development of morality (Ibid, pp.61-66).⁷³ A categorical reason is a reason I have regardless of my personal desires, wants, aims etc. Thus, if I have a categorical reason to keep my promises, I should do so even if I have no desire to. Of course, philosophers sympathetic to *reasons-internalism* will reject this claim.⁷⁴ But the point Joyce is making is that common-sense morality implies that moral reasons are categorical: you should not lie, steal or cheat irrespective of how this fares with your desires. According to Joyce, our conception of the categoricity of moral principles makes evolutionary sense, because it provided us with the psychological motivation to act pro-socially in cases where this conflicts with our personal desires and short-term self-interest. Deciding whether to reciprocate an act of kindness is a lot more straightforward if the organisms believe they have a categorical reason to do so.

Joyce's approach is given support by other evolutionarily inclined normative thinkers. One of whom is Philip Kitcher.⁷⁵ Kitcher, like Joyce, accepts that normativity is the culmination of various factors, some biological, some social. On the biological front, evolution provided us with several essential ingredients, such as altruism, an ability to form commands and obey them and an ability to use language. But, in Kitcher's view, it is only when these are embedded in a social context that ethics arises. In a social context, these capacities are utilised to determine the particular norms which flourish within the society, but their content is contingent on the particular society and the social pressures within it:

Members of the human groups envisaged (small societies, akin to the hominid bands preceding them) are socially imbedded in two important ways. First, as just supposed, the particular way in which normative guidance is internalised depends upon the training regimes present within the group. Second, the content of the orders given depends on discussions among members of the group. The character of the discussions has varied considerably from group to group, time period to time period, with different degrees of involvement according to age, rank and sex. Originally, however, an agreed-on code, articulated and endorsed after discussions

⁷³ The description in the text relates purely to Joyce's conception of a categorical imperative; not necessarily the Kantian understanding of this concept.

⁷⁴ The view that agents do not have external or stance-independent reasons of the kind the realist believes in. In which case, normative reasons derive from some aspect of our psychology, such as our desires, wants, preferences etc. For a classical statement of this view, see Williams (1981, Ch.2).

⁷⁵ Kitcher's approach is best captured in chapters one to four of his 2011 book *The Ethical Project*.

around the campfire, was transmitted to the young through training regimes that had also been socially elaborated and accepted (Kitcher 2011, p.96).

In this excerpt, Kitcher signals his opposition to evolutionary accounts claiming that evolution shaped the content of our normative beliefs, by highlighting the variance in normative beliefs that may arise out of different societies. Kitcher takes a broadly pragmatist approach to morality. Morality is a tool which we use to solve collective action problems. But the particular problems we face are not uniform across cultures, but are very often culturally specific.

How are these content-neutral approaches to evolution's normative influence supposed to undermine realism? It cannot be, as Street claims, because evolution has shaped our normative beliefs in off-track directions, because these philosophers reject this claim. Both Kitcher and Joyce dedicate a chapter of their books to arguing against moral realism.⁷⁶ In some cases, their arguments do not specifically draw upon their evolutionary genealogies. For example, echoing the afore-quoted remark by Street, Kitcher is puzzled about how we came to recognise the realist's objective normative truths, asking: "Where exactly do historical actors bump up against the external constraints and acknowledge their force?" (Kitcher 2011, p.178). This, as I argued in the last chapter, is something of a straw target, at least in the case of non-naturalist, for they do not claim that normative truths are physical things and thus, they are not the kind of things we can bump up against. Kitcher then recites an argument notably made by Mackie (1977, Ch.1), which is that it cannot be the case that we discover normative truths, since the normative beliefs we hold are imparted by our society. That is not to say we cannot revise such beliefs; the existence of social and moral reformers shows we can. But when we do so, it is from within the moral premises we have internalised from our society. And thus, what we are doing is not *discovering* a new normative truth, but showing how our society had inconsistently applied their own principles. However, both Joyce and Kitcher also draw upon arguments relating to their evolutionary genealogies to undermine realism. Kitcher does this in the form of a challenge (Kitcher 2011, pp.186-189). It is a requirement of a metaethical theory, Kitcher claims, that it can be linked up with the correct normative genealogy. Kitcher claims to have provided a plausible genealogy in the form of his evolutionary/sociological thesis and his pragmatist anti-realism is claimed to link neatly with it. Kitcher then goes on to challenge the normative realist to do likewise. I will take up this challenge in chapter 5.

⁷⁶ Joyce (2006, ch.6) and kitcher (2011, ch.5).

The evolutionary debunking component of Joyce's argument is more complicated. He begins by asking us to consider a thought experiment (Joyce 2006, pp.179-180). Suppose that there are such things as belief pills. These pills can cause us to believe a proposition even if it is false. Suppose that a side effect of the pill is amnesia about having taken the pill in the first place and indeed, about the very existence of belief pills. Suppose that I discover, beyond a shadow of a doubt, that my long-held belief that Napoleon lost the battle of Waterloo was caused by my being slipped a belief pill. Joyce thinks that under such circumstances, we would have reason to be sceptical about whether that belief is true, given its problematic cause. Now it might be objected here that this thought experiment does not cohere with Joyce's content-neutral version of the EDA, given that evolution has not, on his view, supplied us with normative beliefs. Therefore, likening evolution's influence to that of the belief pill is misleading. Joyce considers this objection and revises his thought experiment accordingly. He asks us to suppose instead that the fictional belief pills do not "generate particular propositional beliefs but, rather, dispose you to form beliefs involving a particular concept – a concept that would not otherwise figure in your beliefs." (Joyce 2006, p.181). And that, rather than cause you to believe a particular proposition about Napoleon, it has caused you to hold a whole host of beliefs relating to the Napoleon. For Joyce, it is not that the content of any particular moral claim is brought into question by evolution, but rather, the very idea of morality is. Evolution disposed us to form beliefs about a concept – namely, morality – which would not have otherwise entered into our consciousness, and it did so, for the sole reason that this enhanced our reproductive fitness. Thus, to follow the analogy, we have been deceived, like the consumer of the belief pill, to believe in the concept of morality. According to Joyce, this should make us sceptical of the very idea of morality and consequently, any positive moral beliefs we happen to hold.

Conclusion

In this chapter, I have explained the rationale behind the EDA and described the argument in its various forms. First, I showed how some of the worries behind the EDA have antecedents in the early responses to Darwin's theory. In the same period, we also saw the first attempts at devising an evolutionary genealogy; indeed, one could argue that Darwin was the first evolutionary genealogist. I then considered some of the empirical evidence in favour of the evolutionary genealogy, citing work in evolutionary psychology, primatology and game theory. Next, I turned to the normative question of what it would mean if evolution did strongly shape our normative judgments. An optimistic response might be that it poses no problem for

the status of normativity if we can ground normativity in evolution. However, I gave several reasons why this proposal fails; for one thing, it would imply an unappealing normative theory given the negative aspects of evolution; for another, such attempts often seem to illegitimately make a normative inference based on descriptive premises.

I then considered EDA. The basic thought behind debunking arguments generally is that if a belief can be shown to derive from a dubious source, then unless we have further reasons for accepting the belief in question, we should regard it with scepticism. In the case of normative beliefs, evolution should be considered a dubious causal factor because normative knowledge is not something for which evolution would select and thus, if evolution shaped our normative beliefs, it was most likely for reasons unrelated to their truth. A local EDA aims to show a single normative principle was shaped by evolution, but such an argument will only ever have implications in normative ethics for the belief in question. Global EDA take aim at normativity as a whole. This can be done either by showing that the content of a large number of our normative beliefs were shaped by the dubious forces of natural selection or by showing that the very concept of morality or normativity was shaped by such forces. These distinctions are important, for they determine the specific objection against the realist and therefore, should shape the way in which the realist responds. In the next two chapters, I will defend realism against Street's content partial EDA and in the final chapter, against the content neutral account of Joyce and Kitcher.

Chapter 3: Has Evolution Shaped Our Normative Judgments?

Introduction

In the last chapter, I outlined the most influential EDA and divided them into the categories: content-neutral and content-partial. In this chapter, I respond to Sharon Street's content-partial EDA. In section 1, I look at some previous failed attempts to defend realism against this argument. In section 2, I outline what I take to be the best strategies for responding to Street's EDA. The first of these arguments, in section 2.1, seeks to show that even if we accept the basic premises of Street's argument - that is, accept that all, or most, of our normative beliefs evolved - it still would not follow that all versions of realism are undermined. If all normative beliefs are products of evolution, there remains the possibility that we revised such beliefs and formed new ones by more reliable means. If all but a small number of normative beliefs evolved, which seems more in line with what Street actually believes, then it is not clear why the debunking implications extend to the small number of non-evolutionary normative beliefs. In 2.2, I try to identify some of the normative beliefs that elude an evolutionary explanation and argue that realists who hold such beliefs not obviously vulnerable to Street's EDA. In section 3, I outline a revised version of the *argument from disagreement* whose target is anti-realism predicated on a wholesale evolutionary explanation of our normative beliefs. I argue that if normative beliefs did evolve, we should expect a large degree of uniformity in their content. That this does not obtain is a reason for rejecting this version of anti-realism.

1. Previous Realist Responses to the Darwinian Dilemma

1.1. The Argument from First Order Normative Truth (Third Factor Responses)

Let's suppose that I am an ethical egoist considering an EDA against my view. I might argue:

1. If EDA are sound, then an evolutionary explanation of a moral principle undermines its truth.
2. Ethical egoism can be explained by evolution.

Let's not worry about whether (2) is true. Assuming it is, the egoist might claim that:

3. Ethical egoism is true.

Which means that:

4. An evolutionary explanation of egoism does not undermine its truth.
5. Therefore, EDA are unsound.

The egoist is claiming that, in the first instance, we know that egoism is true. Thus, if an argument claims that egoism is undermined by its causal history, that argument must be false - any argument whose conclusion is false is unsound. This line of response to the EDA works by appealing to or assuming the truth of the first-order normative belief being targeted by the debunker. It is referred to in the literature as a *third factor response*.⁷⁷ Critics of this line of argument claim that it amounts to begging the question. With regard to the hypothetical argument for egoism outlined above, the very thing in question is whether egoism is undermined by its causal history. We cannot just assume that ethical egoism is true and thereby reject any proposed debunking argument. Of third factor responses, Street writes:

The general question we are asking ... is ‘Why think that the causes described by our best scientific explanations would have led us to the truth in this domain?’ In answer to this question, it is unsatisfactory to reply, ‘My judgments in this domain are true, and they’re also the ones that the causes described by our best scientific explanations led me to.’ Such a reply offers no reason for thinking that the causes led us to the truth; it merely reasserts that they did (Street 2016, p.321).

But this response seems premature. In some cases, beliefs are so well established that any debunking argument seeking to undermine them must be false. Try swapping ethical egoism with the belief that $2+2=4$ or *I am conscious* and then the line of argument seems plausible. It is just a basic fact that $2+2=4$, which means any debunking argument to the contrary must be unsound. So too, the ethical egoist might say, about the belief that the morally right action is that which maximises one’s self-interest.

It is worth noting that this argument is similar to another kind of argument, one used to justify certain claims *which lack a scientific explanation*. For a while, scientists were unable to explain the flight of bees using aerodynamics. But surely, it would be a mistake to infer from this that

⁷⁷ For an example of such a third factor response, see Enoch (2011). Enoch assumes that survival is somewhat good for us. Consequently, actions which promote survival are also somewhat good. Enoch accepts that, for evolutionary reasons, we would be more inclined to think that those actions are good. In this respect, evolution can actually explain why we form so many correct moral beliefs, for not only is it true that survival and actions which promote survival are good, but it is also in our evolutionary interest to hold these beliefs. However, for similar reasons to those given in the main text, this response seems to assume the truth of the “third factor,” namely, the claim that *survival is somewhat good for us*, and if this claim can be debunked, then the third factor response is stopped in its tracks. For discussions of third factor responses to EDA, see Morton (2019) and Klenk (2018).

bees do not fly. That bees can fly is a well-established fact; if it cannot be explained by aerodynamics, then that's a problem for aerodynamics, not the person who believes that bees fly; just as the lack of an evolutionary explanation for consciousness is a problem for the evolutionary biologist rather than the believer in consciousness. The logic of the two arguments is similar. In this case, the argument is claiming that some fact f 's being well-established means that the lack of a scientific explanation for f is no grounds to deny f . In the case of the debunking argument, the evolutionary explanation for f is supposed to undermine f . But if f is established beyond doubt, then an evolutionary explanation cannot undermine f .

If we are to claim that third factor style arguments are sound in defending beliefs like *I am conscious*, $2+2=4$ and *bees fly* but unsound in the case of *egoism is true* there must be some asymmetry between them. The asymmetry might pertain to how widely these beliefs are held. Just about everyone believes that $2+2=4$ and that they are conscious, but there is significant dispute about the truth of egoism. Perhaps third factor responses work only in the cases of beliefs subject to strong agreement. But even that does not seem to be true. It was once the case that most people believed in the existence of God, but I doubt many would be impressed by a response to a debunking argument aimed at Theism which rejected the argument on the grounds that *it is true that God exists*. Alternatively, we could argue that it is the apparently indubitable status of the belief in $2+2=4$ and *my being conscious* which makes them amenable to third factor responses and not, say, normative and religious beliefs. However, I suspect that that devoutly religious people feel fairly certain that God exists, just as I feel fairly certain that we have an objective reason to avoid pain. That feeling of certainty does not guarantee that the belief's truth cannot be debunked.

In the end, my stance on third factor responses is unfortunately quite messy. I think that some beliefs are so epistemically stable that it is virtually impossible to see how they could be undermined by any debunking explanation. Then I think there is a sliding scale on which, the lower one goes, the less plausible third factor responses become. The third factor response seems sound in the case of *my being conscious*, which seems to me to be necessary given the phenomenological experience I am currently having. $2+2=4$ is on a marginally less secure foundation given the remote possibility of a devious Cartesian Demon or brain in a vat scenario, but still seems virtually incapable of being debunked.⁷⁸ But aside from these marginal cases, I think the argument's plausibility radically diminishes, especially when the debunking

⁷⁸ Descartes (1641/1993) discusses the 'Evil Demon' scenario in the first of his Meditations.

explanation carries significant force. For instance, most of us are very sure and – all things being equal – have very strong reason to believe *that the sun will rise tomorrow*. But it is easy to think of evidence that could abruptly dissuade us of that belief – for instance, if scientific authorities informed us that a meteorite is heading toward earth and will obliterate our planet by tomorrow. We cannot reply: “ah, but you see, we know that the sun will rise tomorrow, so your prediction must be false.” In the case of normative beliefs, I am inclined to agree with Sidgwick that even the most apparently self-evident normative truths are not infallible (Sidgwick 1907, p.211). I am not so confident in my own normative beliefs to believe that it is a conceptual impossibility that they are caused by an illusion. I do not believe they are illusory and one of the aims of this thesis is to demonstrate this. But to simply dismiss the possibility, which the third factor response appears to do, seems to be a sign of overconfidence in the truth of one’s normative beliefs. Therefore, whilst I think that third factor style arguments cannot be dismissed out of hand by the proponent of the EDA, they do not seem particularly convincing when used to defend normative principles.

1.2. Resisting the Massive Coincidence Objection

The massive coincidence objections applies to the realist who asserts that evolution bears no tracking relation to normative truth. Street claims that if the realist goes down this route, thus accepting that the process which shapes our normative judgments is not one that aims at normative truth, then we are left with “the implausible sceptical conclusion that our evaluative judgments are in all likelihood mostly off track” (Street 2006, p.117). It is on this point that Street draws the analogy with the unlikelihood of finding Bermuda with only the help of the wind and tides. But not all philosophers accept Street’s claim that believing in this coincidence involves extreme credulousness. In keeping with the Bermuda analogy, Roger White accepts that prior to setting off on our journey, the prospect of reaching our destination would be slim, “But once we come ashore and see the people and street signs and resorts there can be little doubt that we have arrived at Bermuda (even if we must marvel at our good fortune!). Why can’t we similarly recognize evaluative truth when we stumble upon it even by extraordinary accident?” (White 2010, p.589).

White’s response faces the objection that it is quite plain what evidence could indicate that we are approaching Bermuda. It is less clear what evidence is available to inform us that we have

evolved toward normative truths. Unfortunately, there is no internal truth detector which rings any time we form a true normative belief. Now we might take the analogy very literally. Perhaps if we accept a naturalistic version of realism, we might claim that we have perceptual insight into such evidence, given that normative truths are natural facts. But unlike a fortuitous journey to my desired destination, there are no signs telling me in what direction normative truth can be found. Perhaps the standards of evidence in the case of normativity are of a different kind. But it is telling that White gives us no clue as to what they could be. White does follow the aforementioned quotation with the remark that: “That we have an obligation to care for our children for instance seems about as easy to recognize as anything” (White 2010, p.589). But this is not really *evidence* for the belief in the rightness of kin partiality. If White means that we unquestionably do recognise the rightness of kin partiality, we could accuse him of question-beggingly appealing to the kind of third-factor response discussed previously. But it is hard to see how else to interpret him. Suppose someone is trying to debunk a person’s belief in God by appealing to the origin of their belief. Is it reasonable for them to respond: *that God exists seems about as easy to recognise as anything?* Regardless of whether God exists, this response is just question-begging. But that response is no different to White’s.

2. More Promising Lines of Response

2.1. Does Anti-Realism follow Logically From Street’s Argument?

According to Street, evolution has “thoroughly saturated” our normative judgments with its influence. Let’s for now accept, for argument’s sake, that if evolution shaped a normative judgment, it serves to undermine it. This is a plausible thought given that there is no apparent evolutionary advantage to be derived from knowing true normative beliefs. What follows depends on the extent to which Street thinks evolution shaped our normative judgments. Does she mean that every one of our normative beliefs was shaped by evolution?⁷⁹ I do not think this is what Street means. She never explicitly makes this claim. Moreover, Street concedes that other forces have been salient. She writes: “This is by no means to deny that all kinds of other forces have also shaped the content of our evaluative judgements. No doubt there have been numerous other influences [...]” (Street 2006, p.113). Among which, she includes social, cultural and historical factors, and also accepts that the use of reason plays a role (Ibid, p.114).

⁷⁹ Kahane raises similar questions about Street’s argument (Kahane, 2011, p.118).

After these concessions, Street states her position thus: “My claim is simply that one enormous factor in shaping the content of human values has been the forces of natural selection, such that our system of evaluative judgements is thoroughly saturated with evolutionary influence (Ibid).” But this still does not clarify things. It leaves the following questions unanswered: How many normative beliefs were shaped by evolution rather than the other factors mentioned? Was evolution the dominant force that shaped our normative beliefs? Without more specificity, it is not possible to know exactly what Street’s stance is here. The best way to deal with Street’s argument, then, is to consider several different interpretations to see what follows from each.

An EDA with the Widest Anti-Realist Implications

Let’s suppose that evolution has shaped every single normative belief that has ever been held. This would be the most extreme version of the content-partial EDA. We have the following argument:

1. All normative beliefs were shaped by evolution.
2. Since normative knowledge is not fitness enhancing, evolution is not a reliable guide to forming true normative beliefs.
3. If a belief was caused in an unreliable way, then unless there are other grounds for accepting that belief, we should suspend judgment about whether it is true.
4. There are no independent grounds for holding any of our normative beliefs.
5. Therefore, we should suspend judgment about whether any of our normative beliefs are true.

If this argument is successful, realism is in deep trouble. The argument is valid, but some of the premises seem problematic. As I suggested, even Street does not seem to accept the first premise, a premise to which I will shortly apply some counterexamples. I think the second premise is broadly correct. Though as I attempt to show in the next chapter, there may be a few exceptions. But for now, I will assume it stands up. Many would resist the fourth premise: surely there are some normative beliefs which, even if we can show that they bear a causal relation to evolution, they can be verified independently of evolution. Some reject such after-the-fact attempts as *post hoc rationalisations* (e.g. Haidt 2001). However, there is conceptionally a distinction between post-hoc rationalisations and post-hoc justifications and we cannot simply assume that all such attempts will fall in the former category. I think the anti-realist should concede that it is possible in principle that in the case of some beliefs,

independent justificatory reasons are available, but place the onus on the realist to provide such explanations and then consider them on a case-by-case basis.

The above EDA, in claiming that all normative beliefs are shaped by evolution, is the EDA which has the widest anti-realist implications. But I have suggested the first premise is seriously implausible and there are also objections that can be raised about some of the other premises, some of which I will take up in due course. But let us assume, falsely, that evolution really did shape all normative beliefs and that all such attempts to identify independent justifications involve post-hoc rationalisation. Recall the quotation from the last chapter in which Huxley claimed that, rather than follow the path of evolution in a normative sense, we should do our best to escape it. Even if evolution shaped all our normative beliefs, if we could reject such beliefs and replace them with ones which did not evolve, then presumably these new beliefs would not be subject to the anti-realist implications of the EDA. However, this thought can be resisted on several grounds. First, it could be argued that fallible as we are, it is not possible to shake off the beliefs that evolution gave us. Second, it could be suggested that even if we did shake them off, any beliefs which replaced them would also be influenced by evolution.

The first objection – that we cannot shake off our evolved normative beliefs – is false, and its falsity is known empirically from the fact that some people do reject normative beliefs they once held – and as a matter of necessity, if all normative beliefs evolved then the rejection of any previously held normative beliefs shows an ability to revise them. The second objection concedes that we might do this, but that in so doing, we inevitably replace them with beliefs shaped by evolution. This means that, when people change their mind that abortion is impermissible and come to believe that it is permissible (or visa versa), their initial belief and their subsequent belief were both shaped by evolution, presumably meaning that evolution is responsible for providing people with contradictory normative beliefs (a peculiarity which I will revisit later). There is nothing logically contradictory about such a claim. But one might wonder how the anti-realist can be so confident in it. It cannot be known a priori that all “new” normative beliefs which arise out of a process of revision must be caused by evolution. Can it be established by induction? This is a complicated empirical question that cannot be satisfactorily be answered here. But I do think, as I have claimed, that there are some normative beliefs which are not easily explained in evolutionary terms, and I will therefore resume this discussion when I consider my counterexamples to the claim that normative beliefs evolved.

It follows then, that even if evolution's influence on normativity has been so salient that it has shaped all normative beliefs, this does not rule out in principle the possibility of coming to know normative truths. The reason being that even if our original normative beliefs evolved, it is possible that such beliefs can be revised and new ones formed by non-evolutionary means. Note that this line of argument does not face the massive coincidence objection. The massive coincidence objection applies to those who believe that evolution by chance happened to shape our normative beliefs in the direction of truth. The claim above is that it is conceptually possible that evolution shaped our original normative beliefs in the wrong direction, but owing to our ability to revise evolved beliefs, we were then able to use some other means – perhaps our use of reason – to gain insight into normative truth.

The claim that evolution shaped all normative beliefs is not one that receives much endorsement. Perhaps when Street claims that our normative beliefs are saturated by evolution, she means that all our normative beliefs are somewhat influenced by natural selection, but that evolution is not always the most salient factor. This coheres with Street's suggestion that reason and culture also play a role. The implications of this interpretation depend on the extent to which these respective forces influenced our normative beliefs. If the claim is that, notwithstanding the influence of reason and culture, evolution is always the dominant force shaping our normative beliefs, then the argument that I have just made applies here too. Perhaps Street accepts that in some cases evolution is not the dominant force, and that some normative beliefs were mostly shaped by reason or culture (or some other factor). If so, then whether such beliefs are undermined depends on how well reason or culture track normative truth. I will make an argument in chapter 5 in favour of the view that through the use of reason, we can gain insight into normative truth.

Let's consider another way of interpreting Street. Rather than claiming that all normative beliefs were shaped mostly by evolution, perhaps she means that the overwhelming majority were. But that a small number were not. This could still be bad news for realists if most beliefs are of this kind. Let's suppose, to see what follows, that evolution has shaped every single normative belief except one, $\sim e$. Following Street's logic, the fact that evolution is not a reliable truth tracker means that every single evolved belief - all beliefs except $\sim e$ - is unreliable unless there are independent grounds for supposing them. Suppose there are no such independent grounds. We would then have grounds for taking a sceptical stance toward all normative beliefs except $\sim e$. This leaves the obvious question: what about people who hold that $\sim e$? Although $\sim e$ is a radical outlier in being the only normative belief which was not shaped by evolution,

believers of $\sim e$ do seem to be immune to Street's argument, on the grounds that their belief is not tainted by the influence of evolution. Moreover, if the original premise of Street's argument is that all normative beliefs except $\sim e$ evolved, then the conclusion - *Therefore, we should suspend judgment about whether any of our normative beliefs are true* - no longer follows validly from the premises. To be valid, the conclusion would need to be (5*) *Therefore, we should suspend judgment about whether all of our normative beliefs, except $\sim e$, are true.*

We can see what follows from this in terms of normative ethics. Ethicists who embrace $\sim e$ as a normative principle are immune from EDA. It might be that $\sim e$ can be refuted by some different argument, but not the EDA. The metaethical implications are more complicated and demonstrate that metaethics is not wholly separate from its normative cousin. Suppose that a normative realist holds normative beliefs other than $\sim e$. Since all their beliefs would then fall into the category of beliefs which can be debunked by evolution, her position is vulnerable to global EDA. The evolutionary debunker might then ask this realist to take a stance on whether to assert or deny a relation between evolution and the supposed truth of such beliefs: her answer will either commit her to an unscientific belief in evolution's tendency to track normative truth or belief in an unlikely coincidence. Suppose I am a normative realist that believes that only $\sim e$ is a normative truth. Because Street does her debunking in a bottom-up way, via the claim that evolution shaped our normative beliefs, if I am a realist whose only normative belief cannot be debunked by evolution, my belief in realism cannot be debunked by Street's EDA. This might be considered a hollow victory. If only a single belief is immune from EDA, it would still undermine most realist positions. Indeed, all realists but those who accept $\sim e$ as the single true normative belief will feel the sting of Street's argument. But as Singer and de Lazari-Radek (2014, p.181) point out, such a conclusion might be welcomed by certain realists, particularly those who hold unpopular moral views. They point out that utilitarian realists like Henry Sidgwick could broadly embrace this conclusion, by claiming that all normative beliefs other than the belief that we should maximise happiness are false.

Street could respond by arguing that this entails commitment to another kind of coincidence: how convenient that the one normative belief not caused by evolution just happens to be the one the realist believes in. However, this response rests on the assumption that we would not be able to tell the difference between our evolutionary and non-evolutionary normative beliefs. Perhaps by engaging in some of the speculations found in the evolutionary psychology literature we discover that some normative beliefs simply do not lend themselves to a plausible evolutionary origin story. In which case, it would not matter if there was luck involved in the

realist happening to stumble across the single outlier so long as we could back up the claim that it is amenable to no evolutionary explanation. It follows, then, that even if just one normative belief was not caused by evolution, realists who endorse that belief as the single normative truth would be immune to Street's EDA.

2.2. The Argument from Non-Adaptive Normative Beliefs

Thus far I have been dealing with conceptual possibilities. I have claimed that some realist views could survive in all of the following three scenarios:

- 1) That evolution completely shaped all normative beliefs.
- 2) That evolution partially shaped all normative beliefs.
- 3) That evolution completely or partially shaped all normative beliefs except one.

I think none of the above scenarios describe reality. There is more than one normative belief that cannot be adequately explained by evolution. I do not know exactly how many – indeed, I cannot know, because there are normative beliefs of which I am unaware. The question of how many outliers there are to the evolutionary debunking explanation is a difficult empirical question. Below I posit some candidates.⁸⁰ These remarks are speculative, but no more so than proposed evolutionary explanations. And I concede from the outset that, though no explanation exists for such beliefs presently, one may arise in the future which serves to undermine the argument that follows.

Street's debunking case rests on the claim that our normative beliefs are shaped by evolution. However, some normative beliefs resist evolutionary explanations. Based on the existence of such beliefs, the realist might argue that:

1. In order for a normative belief to be undermined by evolution, it must be the case that the belief in question can be explained by evolution.
2. Some normative beliefs cannot be explained by evolution.
3. Therefore, some normative beliefs cannot be undermined by evolution.

⁸⁰ Appeals to apparently 'non-adaptive' normative and moral beliefs in response to debunking arguments have been pursued by Singer (2005) and Shafer-Landau (2012, pp.7-8).

It is hard to see how anyone can quarrel with the first premise. Surely, if some belief was not caused by evolution, it cannot be undermined by evolution. It would be absurd to cite a faulty clock as undermining my belief that it is 4pm if I did not acquire that belief by reading a faulty clock. Premise (2) is empirical. I will now consider several prospective normative beliefs which cannot be explained easily by natural selection. A caveat is warranted here. Though premise (2) is empirical, it is also speculative (but then so too is its negation). It requires consideration of events that we cannot observe and for which we have no direct evidence. How do we decide whether or not a belief evolved? It is not straightforward, but some of the methods were considered in chapter two. Evolutionary psychologists seek to provide explanations for how some normative belief enhanced the survival and reproduction. Presumably, if no such explanation is available, the belief is not a suitable candidate for an evolutionary explanation.

Non-Adaptive Utilitarian Principles

I asserted at the outset that a defence of utilitarianism, though not my core objective, is one of the aims of this thesis. I think that utilitarians are on particularly strong footing in responding to EDA. However, at this point I am only going to consider three normative principles which are typically, though not exclusively, attributed utilitarians: the *Principle of Universal Benevolence*, the *rejection of the acts and omissions distinction*, and the *commitment to value maximisation*. The reason for not defending utilitarianism generally at this point is that there is significant dispute about whether the hedonistic account of value can be explained by evolution – a question I discuss in the next chapter.

The Principle of Universal Benevolence

This principle is core to utilitarianism generally, though as a principle in its own right, it receives perhaps its most thorough treatment by Sidgwick. According to Sidgwick's *dualism of practical reason*, it is always rational for an agent to act in either of the following ways (Sidgwick 1907, bk.4). She could act according to rational egoism by attempting to maximise her own wellbeing. Or she could act in such a way as to maximise the wellbeing of all, impartially considered. The rationale behind the second option is the principle of universal benevolence, which states that, "[...] the good of any one individual is of no more importance, from the point of view of the universe, than the good of any other; unless, that is, there are some special grounds for believing that more good is likely to be realised in the one case than in the other" (Sidgwick 1907, p.382). The second part of Sidgwick's formulation refers to the permissibility of prioritising the interests of one person over another if that person's benefit is

likely to be greater than the other's. For instance, if I know that giving £1000 to someone in abject poverty will enhance their wellbeing significantly more than someone reasonably well off, then prioritising the person in abject poverty is perfectly consistent with the principle of universal benevolence. What is not permissible is prioritising for reasons which have nothing to do with total wellbeing, such as prioritising my wellbeing *because it's mine*, or someone else's *because they stand in some special relation to me*. Sidgwick struggled to find a way to demonstrate the superiority of the principle of universal benevolence over rational egoism – a task he referred to as “the profoundest problem of ethics”⁸¹ – and ultimately conceded that a rational agent may do either. Though it is worth noting that in their defence of Sidgwick's ethics, Singer and de Lazari-Radek (2014, Ch.7) try to get around the problem by providing an EDA against egoism. Such a vindication works only if no such EDA can be mounted against the principle of universal benevolence. This is the question which I will now consider.

Would believing that the good of all agents should be given equal consideration enhance evolutionary fitness? Taken to its logical conclusion, it means we should assign no more intrinsic weight to our own, our family's, or our closest friends' interests than we do to the interests of strangers (of course, the utilitarian might allow such prioritisation for instrumental reasons).⁸² Consider the following case:

An Inheritance

Having inherited a £1 million, I could use the money to:

1. Maximise my own wellbeing.
2. Maximise my family's wellbeing.
3. Some mix of 1 and 2.
4. Send all the money – aside from whatever I and my family need for bare survival – to effective charities where it will best maximise overall wellbeing.

⁸¹ Sidgwick makes this remark in footnote 296 of *The Methods*.

⁸² As Derek Parfit (1984, ch.1) has argued, if in general having more concern for those we are close to, such as our relatives and friends, produces more happiness than having no such attachments, utilitarians should not, in general, seek to undermine such attachments. But again, such a policy does not necessarily undermine universal benevolence, because the utilitarian would be justifying partiality due to considerations of total wellbeing. If human nature were such that we could advocate complete impartiality in the way we view people's interests without any loss in happiness, then presumably universal benevolence would dictate that we seek to be totally impartial in whom we care about.

The egoist will claim we should choose 1, an option which, for most people, would indirectly support giving a sizeable amount to family and perhaps a contribution to charity. Common-sense would suggest 3 and perhaps some donation to charity. The proponent of universal benevolence should choose 4, because options 1-3 imply that either one's own or one's family's interests carry greater intrinsic weight than the interest of others. If we ought to consider everyone's interests impartially, then we should go with 4 because it has the best consequences when everyone's interests are given equal consideration. Which of the options make evolutionary sense? Based on the picture I painted in the last chapter, human beings have strong altruistic tendencies. Such tendencies are, nonetheless, strongly partial. Virtually nobody thinks we evolved to be pure altruists. Because the evolutionary process is geared toward survival and reproduction, it most likely favoured a blood is thicker than water mentality, lending support to (2) and (3). However, one cannot completely discount the importance of self-interested considerations. Although we did not evolve to be completely egoistic, it is easy to imagine how egoistic tendencies might make sense in some contexts (usually our own survival is a precondition to benefitting kin).

However, humans did evolve altruistic tendencies beyond their immediate family. Such tendencies likely arose to facilitate cooperation and because, in the small hunter-gatherer societies our ancestors inhabited, there was a good chance that any person with whom one interacted might be a relative. Even so, we should still expect our altruism to be graded in such a way that is at odds with the principle of universal benevolence. Most concern, on this model, will be directed to oneself and one's closest relatives. Less, though still strong concern, would be reserved for one's wider family. Weaker altruistic tendencies – though still salient – would be directed toward one's wider group. This is different from the perspective of the universally benevolent person who thinks that we should perceive the interests of all the above as being equal. Moreover, universal benevolence does not end at the level of one's group, but covers all humans and other sentient beings generally. Although we most likely evolved to be altruistic toward our group members, it is less obvious that we evolved to be altruistic to outsiders and members of other species.⁸³

One response here might be to accept that universal benevolence is at odds with evolution, but retort that despite their proclamations, no one is universally benevolent. Perhaps this response

⁸³ Perhaps we evolved to be altruistic toward some other species, such as certain species of canine. But the principle of universal benevolence implies impartial consideration to the interests of all sentient species and thus would not endorse this kind of selective species altruism.

is a red herring. It does not matter if no one is universally benevolent, what is in question is whether some people believe we ought to be. Of course, this response would be undermined by the truth of motivational-internalism, the view that if an agent sincerely accepts a moral principle, she will be motivated to act in accordance with it.⁸⁴ If no one is motivated to be universally benevolent, so this argument goes, nobody really believes we should be. As it happens, I do not find motivational-internalism particularly plausible. I see no reason why someone must be motivated to act upon a moral principle she believes in. Consider the example of an *amoralist*, someone who knows her actions are wrong but does not care. If motivational-internalism is true, the amoralist is a conceptual impossibility (Shafer-Landau 2003, pp.145-157). In any case, I do not believe motivational-internalism undermines the principle of universal benevolence. According to the most plausible accounts of motivational-internalism, for an agent to sincerely hold a moral belief they must have *some* but not *complete* motivation to act upon that belief; that is, an agent can fail to do what she thinks she ought to do due to weakness of the will (Smith 1994, pp.134-135). Although I doubt anyone is perfectly universally benevolent, there are actions which show some motivation in that direction, such as blood donation and impartial organ donation. The fact that no one is perfectly universally benevolent no more undermines the belief in universal benevolence than not being a perfect Christian undermines one's belief in Christianity.

An alternative objection comes from the psychological egoist. Unlike the normative egoist, who claims that we ought to act in a way that best maximises our own interests, the psychological egoist claims that the maximisation of self-interest underlies the way we do act (Shafer-Landau 2018, p.91). This does not mean we never act altruistically. An 'enlightened' strategy for maximising self-interest would mean acting altruistically, given that the wellbeing of others is crucial to our own wellbeing. But, in such cases, the psychological egoist claims that our ultimate rationale is egotistic. The psychological egoist will claim that not only do people not act universally benevolently, they could not possibly do so. However, psychological egoism has some deeply implausible implications. It implies, for one thing, that when people think they are acting altruistically *solely for the sake of others*, that they are under some kind of collective illusion. It requires us, moreover, to be unduly sceptical about the professed motives of altruists, implying the psychological egoist has some kind of privileged access into the mental states of others (perhaps their belief in psychological egoism is a mistaken generalisation based upon their own motivations). But more problematically, it seems very hard

⁸⁴ See Smith (1994, Ch.4).

to justify psychological egoism empirically given that some people act in altruistic ways that are extremely difficult to reconcile with self-interest. The blood donor is no more likely to receive a blood transfusion because she donated, nor is anyone she cares about.

Perhaps the egoist might claim that it is the happiness derived from such acts, on the part of the altruist, which make them ultimately egoistic. It is true that being altruistic can increase our happiness, though whether this is enough to outweigh the sacrifice involved in something like impartial organ donation is highly questionable. And this still leaves the case of those who sacrifice their lives for others, such as Ross McGinnis, a soldier who intentionally put his body against a live grenade to save the lives of his fellow soldiers, sacrificing his life in the process.⁸⁵ As Shafer-Landau (2018, p.93) points out in his discussion of this example, there might be various motivations involved here, and unless we are in that situation, it is impossible to know for sure. But it is hard to believe that McGinnis's action was of self-interest given that it manifestly contradicts what a self-interested person would do under these circumstances. We might say something highly contrived, like McGinnis sacrificed himself because his life would be plagued with guilt otherwise, but this seems to overly rationalise such actions: it is hard to believe that such considerations could have been made in the short time between the grenade falling and his attempt to shield his fellow soldiers. Finally, there is no evolutionary reason for accepting psychological egoism. I showed, in fact, that there are good reasons for believing that altruistic tendencies evolved. Thus, whilst I accept that people sometimes act on the basis of self-interest, and indeed, that apparently altruistic acts sometimes disguise selfish motives, there is no empirical reason for believing all acts are egoistic; insisting otherwise seems more of an article of faith than a belief grounded in evidence.

Beliefs that Follow from The Principle of Universal Benevolence

There are other normative beliefs related to universal benevolence which can be taken as stand-alone principles. A belief in universal benevolence is not a pre-requisite for such beliefs, but can be cited in support of them. One is the belief in equality irrespective of characteristics such as one's birthplace, species, race, gender or sexual orientation. To clarify, I do not think that *being prejudiced*, as such, is an evolved trait, but due to early humans inhabiting small groups - groups which competed with other groups for resources - a kind of "parochial altruism" may have evolved, marked by altruistic tendencies that largely fizzle out beyond the level of the group. Based on such considerations, Joshua Greene makes a persuasive case that evolution

⁸⁵ This event is told by Kelly Kennedy in *They Fought for Each Other* (2010, p.108-108).

might have played a role in the development of an “Us versus Them mentality” (Green 2014, pp.48-55). This kind of mentality manifests itself in *racist* and *nationalist* ideation. It is important to make two clarifications. In claiming that such biases have an evolutionary basis is not to claim that they are in any way inevitable,⁸⁶ nor that environment and culture plays no role. Second, Greene is not claiming that human beings evolved to be racists or nationalists. Indeed, claiming that racism is hardwired makes very little sense given that “In the world of our hunter-gatherer ancestors, one was unlikely to encounter someone whom, today, we would classify as a member of a different race” (Greene 2014, p.54). What evolved, then, was the tendency to seek out markers as to a person’s ‘group identity’ to ensure that altruism favours members of *my group*. These markers vary depending on societal and cultural contexts - and indeed, can be completely arbitrary - but modern examples include, and are by no means limited to, *race*, *nationality* and *names* (e.g., name bias in job applications).

The principle of universal benevolence asserts that we should give equal consideration (though not necessarily equal treatment⁸⁷) to the interests of all species and therefore to not prioritise the interests of our own species – except for reasons that pertain to overall wellbeing as opposed to the morally neutral fact of species membership. That some pain or pleasure is experienced by a human is not, in and of itself, a reason for giving it priority over other species’ pleasures and pains.⁸⁸ Some have argued (notably Singer⁸⁹) that this principle tends toward ethical vegetarianism or veganism. I find it extremely difficult to see how a belief in the rightness of ethical veganism can be claimed to be an evolved moral principle. The view is fairly recent in human history and is held by a relatively small number of humans.⁹⁰ For another, there is a good evolutionary explanation for *Carnism* – the view that it is permissible to consume meat. Namely, that for our pre-agricultural Stone Age ancestors, hunting animals for meat consumption was integral to survival.

Peter Singer’s *expanding circle thesis* claims that over time, human beings have gradually widened their sphere of moral concern (Singer 2011). Starting from the level of the group, we have come to widen our concern to our wider society, humanity as a whole, and to other species.

⁸⁶ Nor remotely justifiable. For reasons outlined in chapter 1, there is no reason to think that some trait’s having evolved serves to justify it.

⁸⁷ Having given equal consideration to the interests of two species, we might decide, purely on the basis of the amount of wellbeing they are capable of enjoying, that one species should be given priority over the other. And of course, equal treatment would be pointless in cases where organisms have different interests.

⁸⁸ Singer refers to this bias as *speciesism*.

⁸⁹ For a discussion on this, consult the first chapter of *Animal Liberation* (1975/2015).

⁹⁰ I mean recent in terms of overall human history. Ethical vegetarianism has a rich legacy in non-Western religious traditions such as Hinduism, Buddhism and Jainism.

The proponent of universal benevolence should also claim that our sphere of concern should be expanded temporally, to assign equal intrinsic weight to the interests of future people. I use the term *intrinsic* because this principle is commensurable with discounting future interests for reasons of uncertainty. But importantly, it means we cannot discount *just because* the interests in question occur in the future. Sidgwick considers *temporal neutrality* (or the *axiom of prudence*) as an axiom in and of itself (Sidgwick 1907, Bk.3, Ch.13); and it can be applied egoistically (I should not prefer a pain to happen in the future *just because* it happens in the future). But when the principle is applied interpersonally, it implies equal consideration of the interests of sentient beings irrespective of the time at which those beings exist.

It seems unlikely that we evolved to be impartial about the interests of future and present generations. The way humans typically behave suggests that we prioritise the short-term. This is often the case with self-directed behaviour, such as when children forgo a more desirable treat in the future for a less desirable one in the present. Of course, humans are capable of acting prudently. But when we do so, it often feels as though our rational foresight is prevailing over our more instinctive motivations. When it comes to other-directed concern, there is no question that human beings are naturally short-termists. That much is clear from the excruciating lack of progress we have made in changing our behaviour to mitigate the effects of anthropogenic climate change, whose most tragic consequences will be felt by future generations. There are obvious evolutionary reasons for why we think short term (Martin Daly & Margo Wilson 2005, pp.55-60). The evolutionary pressures our ancestors faced were not of the kind which required calculations spanning billions of people over hundreds and thousands of years. Such considerations are very recent. Of course, many people do make radical sacrifices in response to climate change, responses that would be difficult to rationalise in terms of enlightened self-interest, kin partiality or reciprocal altruism. An example highly relevant to evolutionary psychology being the choice some people make to refrain from having children due to worries about overpopulation and its impact on climate change. But this is an argument against the global EDA, rather than for it, because it means that many people believe and act upon moral principles that, if anything, conflict with our evolved normative mentality.

Rejection of the Acts and Omissions Distinction

According to utilitarians, there is no direct moral difference between acts and omissions (see Singer 1993, Ch.7). For instance, if I fail to act to prevent a death, this is no better or worse than actively taking a life, unless there is some other morally relevant difference involved to

explain the asymmetry. But the mere fact that one death was caused by an action rather than inaction, cannot make a moral difference in the way we evaluate such cases. This conflicts with the widely accepted acts and omissions distinction, according to which, all things being equal, it is worse to bring about a harm (such as a deliberate killing) than to omit to prevent one (such as not donating to life saving charities). Again, the question arises: what stance would we take on the acts and omissions distinction if our belief on the matter was shaped by evolution? The first thing to point out, which applies also to universal benevolence, is that the person who rejects the acts and omissions distinction is very much in the minority. Her belief plainly conflicts with the dictates of common-sense morality, which endorses various kinds of partiality and strongly endorses the acts and omissions distinction. This is relevant because if a trait or attribute is evolved rather than caused by culture or something else, we should expect it to be fairly universal across our species. Because common-sense beliefs by definition are those which tend to be more universal, these are the ones best suited to an evolutionary explanation. But if the belief in question – like universal benevolence and the rejection of the acts and omissions distinction – are held by a relatively small number of people, they are less likely to have evolved.

Roger Crisp has proposed an evolutionary explanation for belief in the acts and omissions distinction. He asks us to consider the following example:

Blindness. On Monday I blind a stranger to prevent his buying the last copy of a CD I want to buy. I buy the CD. On Tuesday I buy another CD, knowing that I could have given the money to Sight Savers International and prevented the blindness of at least one person (Crisp 2006, p.21).

As Crisp points out, common-sense forbids the act of blinding the stranger, but seems to permit my purchasing the CD with money that could have been used to prevent the blinding. Notably, the consequences of the two actions are equal (in terms of deaths resulting from the act or omission). According to Crisp:

This kind of morality [that actively killing is worse than letting die] is what we would expect to have emerged from the evolutionary process. It is clear that a group cannot function well if its members are permitted to harm one another, whereas the survival value of a prohibition on allowing others to suffer is more dubious (Ibid.).

The utilitarian could plausibly argue that Street is right to a point. Here we have an example of a belief – that there is a moral asymmetry between acts and omissions – that can be explained

by evolution. And in being amenable to this explanation, we should be sceptical about the belief. But evolution cannot explain the fact that some, such as utilitarians, hold the negation of that belief, because that belief is contrary to what we should expect evolution to produce.

Value Maximisation

The third respect in which utilitarianism conflicts with common-sense morality concerns the duty to maximise. Though the two are conflated, this principle differs from universal benevolence. Universal benevolence just states that we should give everyone's interests equal consideration and that any discrepancy between how we treat *a* and *b* can only justifiably be based on some morally relevant factor, such as the fact that *a* is capable of more happiness, or the probability of successfully bringing about happiness in *b* is stronger. This implies nothing about a general duty to maximise wellbeing. I also think that the two principles arise out of different intuitions. That we should maximise wellbeing seems not to emanate out of considerations of equality, but from the basic thought that, all things being equal, if world *x* contains more value than world *y*, we have sufficient reason to bring about *x*. Taken to its logical conclusion, the best possible world is the one containing the most value. In light of this, a plausible ethical principle would be to act to maximise value, hence the principle of maximisation. Of course, many argue that just because it is good to produce value, there is no general duty to maximise it.⁹¹ But the onus is then on the one who accepts a satisficing account of utilitarianism to draw a non-arbitrary line specifying when one has gone above what is necessary. If no such line can be drawn, parsimony dictates that we should either say we have no duty whatever to maximise value or that we have a duty to maximise value; the more plausible option, I would suggest, is the latter.

According to act utilitarians, it is not just that we ought to do our fair share in maximising happiness, we ought, in every given act, to try to maximise total happiness. It is unclear exactly what common-sense expects of us. I suspect Ross is on the right lines in claiming that common-sense morality provides us with a general duty to promote the good except where this leads to the infringement of rights, or prevents us from fulfilling other duties, such as keeping promises (Ross 1930/2002, Ch.2). Intuitively, I think common-sense also seems more sympathetic to the idea of *supererogation* (Archer 2018). We all have some duty to promote the welfare of others,

⁹¹ A view known as *Satisficing Consequentialism*. See: Slote (1984).

and though the limits to this are vague, *there definitely are limits*. Those who make enormous sacrifices for the welfare of others are not fulfilling their duty, they are going beyond the call of duty. And though their actions should be lauded, they should not be deemed obligatory, which is what the act utilitarian believes. It is also what evolution should lead us to believe. A tendency to believe that each should do their fair share and that free riders should be punished seem like plausible preconditions to the kind of cooperative activities which enhanced the survival and reproduction of our ancestors. It is less clear how acting to maximise the wellbeing of others makes evolutionary sense. Especially when we consider the logical implications of this principle, which suggest that the right action is to make massive personal sacrifices for others, most of whom bear no relation to us.

I believe that this discussion goes some way in defending realism plus a commitment to utilitarianism against EDA. This defence will be completed in the next chapter when I consider whether the hedonistic account of value can be defended against EDA. But I do not think utilitarian principles are unique. I will now briefly summarise some other contenders – some, perhaps all, can be accepted by utilitarians as being instrumentally right.

Choosing not to Reproduce

Evolution is geared toward genetic survival and reproduction. In some cases, this does not lead to the organism themselves trying to reproduce; in the case of bees, it can lead to behaviour which promotes a relative's reproduction. Nevertheless, given the nature of evolution, it is not unreasonable to assume that if our normative beliefs did evolve, we should expect strong normative approval of actions which tend to lead to reproduction. But this is not always so. The use of contraception is an example of people deliberately choosing not to have offspring. Dawkins goes as far to assert that “every time we use contraception we demonstrate that brains can thwart Darwinian designs” (Dawkins, 2003, p.11). Of course, that someone uses contraception does not mean that they never intend to have children. But it is worth bearing in mind that evolution is not just geared toward reproduction, but to maximising reproduction. But aside from that, we have the more difficult task of explaining some peoples' choice to never have children, such as Catholic Priests, Nuns, and those who believe we should not have children due to concerns about climate change, such as Dale Jamieson, who writes: “[...] one moral ideal that I share with many environmentalists is voluntary childlessness (or that, at most, people should have only one child). In almost every case acting on this principle is to act against one's own biological fitness” (Jamieson 2002, pp.323-324).

Rejection of Retributivism

Retributivism is the idea that wrongdoers *deserve* punishment. Not because punishment will produce good consequences (though this may be a supplementary reason), but because wrongdoings make the wrongdoer intrinsically deserving of punishment. On some retributivist views, the punishment the wrongdoer deserves should involve some suffering or pain; on others, that their lives go less well (see Berman 2013, pp.89-88). It is hard to doubt that human beings have retributive tendencies. If someone wrongs me, there is a high chance I will have some desire that they get their comeuppance. Of course, there are good consequentialist reasons for why we might want to punish wrongdoers which do not presuppose the idea of desert.⁹² But the notions of desert and retribution, irrespective of consequentialist justifications, clearly play a strong role in human attitudes toward punishment in their own right. Such attitudes make evolutionary sense. Cooperation works best when each member of the group pulls their weight and therefore a hostile attitude toward free riding might have developed. Anti-social actions, such as theft and arbitrary violence disrupt social harmony, meaning a vengeful attitude toward wrongdoers might have helped keep such behaviour in check. I discussed earlier how tit-for-tat is a strategy forms the basis of reciprocal altruism. This not only encompasses benefits put also harms. If *a* harms *b*, there is a good chance *b* will be inclined to take revenge if presented with an opportunity, if for no other reason than to deter others from acting in a similar way.

Human nature might be somewhat retributivist, but people have come to reject retributivism from a moral perspective. In Plato's dialogue *Crito*, Socrates and Crito consider whether it is ever reasonable to wrong another person.⁹³ "Certainly not, Socrates" answers Crito. Socrates considers a counterexample: "Well, then, is it right to requite evil with evil, as the world says it is, or not right?" Crito claims that such a retaliation would still be wrong, a claim Socrates accepts. Socrates then urges caution on the part of Crito to not accept this claim unless he truly believes it, noting that there are "few who believe or ever will believe this." I am inclined to agree with Socrates that few do believe this. I have no conclusive evidence, but anecdotally, there seems to be widespread agreement that people who commit heinous acts intrinsically deserve to be punished. The prevalence of this belief is not particular to any one culture or time. Jesus referred to it when he said: "You have heard that it was said, 'An eye for an eye and a tooth for a tooth.' But I say to you, Do not resist the one who is evil. But if anyone slaps you on the right cheek, turn to him the other also" (Matthew 5:38-41, *English Standard*

⁹² For instance, to rehabilitate the wrongdoer, to protect society and to deter other would-be wrongdoers.

⁹³ The quotations are taken from pg. 173 of H.N Fowler's 1914 Translation.

Version). In the second volume of *On What Matters*, Derek Parfit expresses his dismay about the strong disagreements between himself and other philosophers. Echoing Mackie's *Argument from Disagreement* (Mackie 1977, pp.36-38), he worried that such disagreements have the potential to undermine his belief in moral realism. One such disagreement relates to retribution. In Parfit's view, no one *deserves* to suffer. Parfit notes that many philosophers disagree with this, but claims that entrenched disagreement about some proposition might not undermine objectivism if it turns out that one side's view is distorted in some respect. Parfit suggests, in line with what I have just claimed, that this retributivist perspective some of his fellow philosophers find appealing can be explained by its evolutionary utility (Parfit 2011b, p.429).

Anonymous Charitable Giving

I suggested that altruistic acts, especially when directed toward strangers, from whom there is no sensible expectation of reciprocity, do not make obvious evolutionary sense. Although there might be no direct advantage to be derived, perhaps there are indirect benefits. Some evolutionary psychologists have suggested that reputational enhancement can provide an explanation for altruism (Greene 2014, pp.58-61). Those who attain the reputation of being kind may derive various social advantages. But in order for this to happen, one's altruism must be public knowledge, or at least, be known by someone from whom I can expect to derive some later benefit. Often this is the case, such as when large corporations donate money to some cause. Without being too cynical, it is hard to believe that positive publicity is not part of the rationale behind such altruism. Of course, there might be good reasons for publicising charitable acts which have nothing to do with reputation (Singer 2009, Ch.4). As some effective altruists have pointed out, practicing what one preaches might lend credibility to a charitable cause, and the more people who publicly engage in charity, the more pressure that is exerted on those who do not. But some people have argued that the moral worth of charitable acts is contingent on their being kept secret. For instance, Jesus warned against "[...] practicing your righteousness before other people in order to be seen by them [...]" and that "[...] when you give to the needy, sound no trumpet before you [...] so that your giving may be in secret" (Matthew 6:1-4, *English Standard Version*).

This is by no means an exhaustive list of all the normative judgments which seem to elude an evolutionary explanation. But they are enough to illustrate my point. Namely, that the evolutionary explanation of our normative beliefs is limited. Some normative beliefs seem to

conflict with the ‘aims’ of evolution. Which means that realists who hold such views are not necessarily vulnerable to the content partial EDA. I will now consider some potential worries.

2.3. An Objection: Is this Argument a Red Herring?

It could be argued that my attempt to show that some normative beliefs conflict with evolution is a red herring and does not in any way enhance the credibility of realism. Someone could argue:

Evolutionary psychologists claim there is a good evolutionary explanation for why people like sweet foods. For our ancestors, fruit was an essential source of nutrition, so it would make sense if humans evolved preferences for sweetness. However, some people do not desire sweet things. Therefore, humans did not evolve preferences for sweet foods.

Another example:

Biologists claim that vision evolved to enable people to navigate their surroundings. Some people are blind. Therefore, vision did not evolve.

Clearly, such arguments are absurd. However, someone might argue that my positing people with beliefs that did not evolve as counterexamples to an evolutionary account of norms is equally flawed. There is a subtle difference between these two proposed analogies so I will treat them separately. But in either case, despite surface appearances, these arguments differ from the one I have made.

In the second case, the explanation for blindness can generally be attributed to a biological dysfunction. The biological function of the eye is vision. If a person is blind, their eyes are, in an evolutionary sense, not functioning correctly, which means that blind people are not counterexamples to the claim that vision is something which evolved (all it shows is that evolutionary traits sometimes do not serve their ‘evolutionary purpose’). But it is for this reason that this example is not analogous to the claim that I have made, unless we want to say that belief in universal benevolence or the wrongness of retributivism involve an evolutionary dysfunction - e.g., that the proper function of normativity, which is to produce kin-partialist and pro-retributivist tendencies is not working correctly. Assuming that acceptance of the views I have mentioned does not involve an evolutionary dysfunction, we cannot claim that the second argument is analogous to the one I made.

Case 1 seems more pressing. I suppose, at a push, we might want to say that not liking sweet foods is a kind of evolutionary dysfunction, given that this preference might have been harmful for our ancestors (though of course, from our modern perspective, it might be a blessing given the harmful consequences of consuming excessive amounts of sugar). In any case, I think there is another relevant difference. The claim that we evolved to like sweet tasting food is different from the claim that we evolved to enjoy all sweet foods. Just as the claim that we evolved to hold certain normative beliefs is different from the claim that we evolved to hold all our normative beliefs. If a person does not like the sweet taste of candy, then they serve as a counterexample to the claim that we evolved to enjoy all sweet foods, but not to the more modest claim that we evolved a general liking for sweet foods. Likewise, the fact that some people hold normative beliefs that cannot be explained by evolution is a counterexample to the claim that all normative beliefs evolved, though not to the claim that some normative beliefs evolved. As it happens, I think evolution can explain a lot of our normative beliefs. I also think that it can explain our general liking for sweet foods. But these are big picture claims, whereas my remarks are more limited. I am merely suggesting that there are limits to explanations such as these and that – quite specifically – *some* or *a few* normative beliefs are not best explained by evolution. What I am arguing against is the wholesale reduction of normative beliefs to evolution.

3. An Argument from Disagreement

The argument from disagreement is generally considered an argument against realism. It is an argument often attributed to John Mackie (1977). I think there is a version of the argument which undermines content-partial evolutionary genealogies of ethics. I will start by outlining Mackie's original argument, before constructing my revised version. Like all moral realists, Mackie was a cognitivist (1977, p.35). Unlike all moral realists, Mackie denied the existence of stance-independent moral facts. Mackie accepted that moral judgments express beliefs but claimed such beliefs are uniformly false (1977, pp.15-18). He was, therefore, a moral sceptic. Mackie's argument begins with the following premise:

1. There is radical disagreement concerning moral principles.

Mackie discusses three types of moral disagreement, these are: disagreement between cultures, intergenerational moral disagreement, and “differences in moral beliefs between different groups and classes within a complex community” (1977, p.36). These can be summarised as:

a) *Intercultural Moral Disagreement*: When the preponderant moral beliefs of some culture or community c^1 are in conflict with the preponderant moral beliefs of some different community c^2

b) *Intracultural Moral Disagreement*: When the beliefs of some people or group(s) within a culture or community are in conflict with the moral beliefs of other people or group(s) within the same culture or community

c) *Intergenerational Moral Disagreement*: When the moral beliefs of people existing at some time t^1 are in conflict with the moral beliefs of people existing at some past or future time t^2

A further kind of moral disagreement which Mackie does not discuss is:

d) *Intrapersonal Moral Disagreement*: When the moral beliefs of a person at some time t^1 conflict with the moral beliefs held by the very same person at some different time t^2

Let me clarify what this final kind of disagreement entails. People change their views over time. I might accept a moral principle at some point in my life only to revise it or reject it later. If at time t some person p accepts some moral proposition mp , then at some later time t^2 the same person comes to believe that $\sim mp$, the correct description is that since p remains the same person from t to t^2 , she disagrees with her earlier self. To insist otherwise just because it occurs within rather than between lives seems arbitrary. There is another kind of intrapersonal disagreement which is less amenable to a one-line definition. I am thinking here of the kind of in the moment disagreement we have with ourselves when we are said to be in two minds over some moral issue (Coliva 2019, p.7). Suppose that I am trying to determine my position on some moral issue, for instance, on how I ought to act given the fact that global temperatures are rising due to human behaviour. My thought process may lead me in several directions and a kind of internal debate might ensue in which I am attracted to several opposing propositions:

‘I know future generations will be adversely affected, but my individual contribution counts for virtually nothing,’ ‘but what if everyone thought like that?’ and so forth. Although the outcome of such a thought process will be one’s coming down on one side of a question, during the time in which I am forming my view it could be said that there is a disagreement going on within my own mind.

With the relevant forms of disagreement outlined, we can now move to Mackie’s second premise. Mackie claims that moral disagreement has metaethical relevance, writing that “radical differences between first order moral judgments make it difficult to treat those judgments as apprehensions of objective truths” (1977, p.36). This implies that:

2. If there is radical disagreement concerning moral principles, then it is doubtful that any such principles are objectively true

If we accept (1) and (2) then we are forced to accept Mackie’s conclusion that:

3. It is doubtful that any moral principles are objectively true

The first premise (1) is empirical; one that we should by and large accept. Although there is a small number of moral beliefs that might be said to be universal, such beliefs do not constitute the majority. Mackie only claimed that moral disagreement is radical, not absolute. Moreover, the realist cannot explain moral disagreement by attributing it to public ignorance of moral philosophy, for this defence implies that such disagreement dissipates at the academic level, which it does not.

As Mackie points out, the realist’s situation would be less problematic if she could explain moral disagreement (1977, p.36). That is, if moral facts figure in the best explanation of our moral discourse, of which disagreement is a feature, then realism is tenable. But, if there is a leaner explanation which omits moral facts, then it is to be preferred on the grounds of parsimony. Mackie’s argument, therefore, is an inference to the best explanation. As it happens, Mackie denies that moral facts carry explanatory weight when it comes to moral disagreement. According to him, this phenomenon is best explained by social conditioning and participation

in different ways of living. If, to take Mackie's example, one is raised in a society in which monogamy is practiced, then one is likely to be a monogamist. If one is socialised in a polygamist society, one is likely to endorse polygamy. To account for the disagreement between monogamists and polygamists, we need not claim that there is an objectively optimal number of spouses; but rather, that beliefs on the subject are socially conditioned. If Mackie is right, we should accept (2) and therefore (3). I am not convinced that we should accept (2) however, since I believe that the realist can explain the various types of moral disagreement mentioned above in a plausible way. This thesis will be argued for later. For now, I will park Mackie's argument from disagreement, though I will offer a response to it at the end of this chapter. My revised argument from disagreement is as follows:

1. Early anatomically modern human beings evolved in response to a particular set of environmental pressures present during the Pleistocene.
2. As descendants of early humans, the biological functions of all subsequent human beings, including the current population of humans, were shaped by the same environmental pressures.
3. Since all modern humans descended from ancestors responding to the same environmental pressures, we all came to share basically the same set of biological functions, including cognitive functions (a universal human nature).
4. According to evolutionary ethicists like Sharon Street, one evolved cognitive function is a disposition to hold certain moral attitudes because these attitudes promoted survival and reproduction during the Pleistocene; our current set of moral beliefs are mostly by-products of these ancient moral attitudes.
5. If our moral beliefs mostly arose out of the ancient attitudes shaped by environmental pressures present during the Pleistocene, then given that these environmental pressures were universal in the way they shaped human cognition, we should expect the emergent moral beliefs to be roughly the same across out species.
6. However, it is not the case that moral beliefs are roughly the same across our species, since there is extreme moral disagreement.
7. Therefore, it is not the case that human moral beliefs mostly evolved in response to Pleistocene environmental pressures.

I will now summarise the premises. (1) is an empirical premise about the evolution of modern humans endorsed by many evolutionary psychologists. From this premise it logically follows that the ancestors of all living anatomically modern humans, such as ourselves, were shaped, both in terms of their physical and mental characteristics, by the same set of environmental pressures (2). As a result, all humans came to share the same biological and cognitive functions (3). That is to say, notwithstanding minor variations - mostly in the molecule-by-molecule sequence of our proteins (Pinker 1997, p.49) - there is a universally evolved human nature. Tooby and Cosmides (1992, p.78) write:

Why do we so often connect complex adaptations or evolved architectures with concepts such as *species-typical*, *human universal*, *species-standard*, *recurrent* and so on? This is because when humans are described from the point of view of their complex adaptations, differences tend to disappear, and a universal architecture stands out in stark relief. (emphasis in the original text).

Street bases her moral genealogy on the work of evolutionary psychologists and presumably, therefore, would endorse the premises asserted so far. Street claims that the evolutionary pressures to which our ancestors responded shaped the content of most of our moral judgments (4). However, since our ancestors were responding to the same environmental pressures, we should expect, as with other cognitive functions, that these moral judgments were shaped in the same direction. And if our moral judgments were mostly shaped in the same direction, we should expect strong moral agreement (5). But, pace Mackie, our moral attitudes are far from the same (6). Empirically, we know extreme moral disagreement obtains and therefore, it is not the case that moral beliefs were shaped in the same direction. Again, in making this claim, I am not suggesting no moral beliefs are universally held by humans. However, Street's argument is not that some moral beliefs were shaped by evolution but that most of them were. The consequence of this would be, assuming evolution deterministically shapes our moral beliefs, that most of these beliefs turned out to be the same. This clearly did not happen and therefore we are left with the conclusion (7) that our moral beliefs are not mostly products of evolution. Finally, since (7) negates the foundational premise on which Street's evolutionary debunking argument rests, we should reject that argument.

It is worth stepping back for a moment to consider why this argument specifically targets Street's content-partial EDA and not content-neutral versions. Joyce and Kitcher think there is

an evolved ability to engage in ethics but deny this extends to first order moral beliefs. Since they attribute the content of our moral beliefs to several factors, there is no reason to expect moral uniformity. Moreover, accounts like Mackie's, which give priority to culture also do not imply uniformity, since cultural pressures vary strongly between and within human societies. Street's content-partial evolutionary account, it would seem, is unique among anti-realist views in implying that human moral beliefs should strongly converge.

Street might deny this by pointing out that it is possible for conflicting beliefs to arise out of a shared cause. For instance, experience of the brutality of war can shape both pacifistic and militaristic moral viewpoints. Therefore, Street might deny that evolution would have shaped our moral beliefs in the same direction. This is a plausible response, but only if we reject Street's content-partial evolutionary account, for this account suggests that evolution's influence on human morality was largely deterministic, since Street claims that given the way that we evolved, it is no accident that our moral beliefs turned out exactly as they did (2006, pp.115-116). Given that we are social animals who rely on cooperating with others, it is no accident that we developed pro-social moral beliefs. Given that the unit of selection in evolution is the gene, it is no accident that people think they have stronger moral obligations toward family members than to strangers. Thus, certain evolutionary pressures lead to particular moral beliefs, and since all our ancestors were responding to very similar evolutionary pressures, it is reasonable to expect that these beliefs turned out the same.

3.1. Explaining Moral Disagreement

Following Mackie's lead, I consider this argument to be an inference to the best explanation. My argument provides *prima facie* grounds for rejecting Street's position, which might be overcome if an adequate explanatory thesis is brought to bear. I will now consider whether such a thesis is available. I will argue that the answer is no, and that the realist's explanation of moral disagreement is more plausible. This conclusion suggests both that (1) Street's evolutionary debunking argument against realism fails because its first premise is false and (2) the phenomenon of moral disagreement is more of an issue for Street's anti-realism than it is for realism.

3.2. Intercultural and Intracultural Moral Disagreement

My argument does not deny that evolution can explain some moral beliefs. I claimed, rather, that evolution cannot explain most of our moral beliefs. I think that the existence of moral disagreement bears this out. That is, I think content-partial evolutionary accounts of ethics struggle to explain scenarios in which some people accept some moral proposition *mp* and others the negation of that proposition $\sim mp$. There is much disagreement within cultures over issues ranging from abortion, euthanasia, gay marriage, and whether or not taxes are too high. For simplicity, let's consider one of these issues. In the USA, there are many people who accept the moral proposition that:

mpa: Abortion is morally permissible.

There are also many who deny that proposition, they believe that:

$\sim mpa$: Abortion is morally impermissible.

On Street's account, our views concerning the permissibility of abortion are presumably explained by the evolutionary forces acting upon our ancestors. But these forces were largely uniform across our species. It is not the case that people who are pro-life evolved differently to people who are pro-choice. Perhaps evolution just selected for one of these moral beliefs; for argument's sake, the belief that abortion is impermissible. Even if this were so, we would still be lacking an explanation for why some people reject that belief (and visa versa if evolution favoured *mpa*). In either case, it must be that at least one of these beliefs was formed by non-evolutionary forces and therefore, the evolutionary explanation of these beliefs will be limited. A final possibility is that evolution selected for neither belief concerning abortion, in which case, the evolutionary account utterly fails as an explanatory thesis. The same applies to all moral beliefs on which there is substantial disagreement. At best, the evolutionary explanation will only explain one side of the argument (either *mpa* or $\sim mpa$); at worst, it will explain neither side, but it cannot be that evolution explains both sides, and therefore, the claim that evolution shaped most of our moral beliefs is undermined. Similar things can be said about cross-cultural

moral disagreement. If one culture contains more proponents of *mpa* than another, it is not because that culture evolved differently: non-evolutionary factors must be posited to explain the disagreement.

3.3. Intergenerational Disagreement

Many societies have shifted in their moral beliefs over time. This applies to views regarding the permissibility of slavery, absolute monarchy, capital punishment, poverty, torture, sexism, racism and homophobia. In some societies, views which were once widely held no longer are. Some of these shifts were more abrupt (and decisive) than others. Others more gradual and limited. For now, we need not assume any of these changes were improvements, only that they are changes. The question is how to explain them. The problems Street faced in accounting for intercultural moral disagreement apply here too. In the above examples, the majority of people comprising some society at time t believed some moral proposition *mp* whereas at some later time t^2 the people within that society accepted $\sim mp$. Since the people comprising these societies did not evolve differently, evolution cannot explain one society's belief that *mp* and another's that $\sim mp$. Perhaps, in some of these cases, evolutionary forces played a causal role on one side of the disagreement. Suppose, for instance, that an evolutionary in-group bias can explain the prevalence of racist views in some society. Even if this were so, it would not explain how members of the same society came to reject such views. Furthermore, evolutionary debunkers are ill-equipped to deal with what Michael Huemer has called the "recency" of some of these changes (2016, p.1995). Lasting evolutionary change takes a long time to occur, often many thousands of years. Some of these changes occurred within centuries; some within decades. It is not plausible to claim that these changes happened because moral principles which were once adaptive suddenly became maladaptive. To quote Huemer: "there is no reason to think, for example, that in the 1960's racists started having fewer children than non-racists and thus failed to pass on their racist genes, or that during the last 200 years, people who supported democracy started having more children than those who supported dictatorship" (Huemer 2016, p.1995).

3.4. Intrapersonal Disagreement

Intrapersonal moral disagreement involves some person changing their views over the course of her life, thereby disagreeing with herself over time. At one point in a person's life, they might accept *mp* and at a later point $\sim mp$. Suppose that as a young man, Frank was committed to left wing ideals but as he aged, he became increasingly conservative. The following argument can be made:

1. Both young Frank and old Frank were one and the same person
2. Both young Frank and old Frank are genetically identical.
3. If moral beliefs are genetically determined, then young and old Frank would hold the same moral beliefs.
4. Young and old Frank did not hold the same moral beliefs
5. Therefore, moral beliefs are not genetically determined.

This argument can be resisted on several grounds. The most objectionable premise is (3). One might object that (3) does not reflect the views of any evolutionary ethicists, who claim that the phenotypic expression of a gene is always dependent on environmental factors (van Vugt 2014, p.7) For instance, although I might be preconditioned to be partial to the interests of my family members, if I were maltreated by some relative, this environmental stimulus might suppress my initial instinct to be kind to them. However, people sometimes change their moral views without any discernible change in their environmental surroundings. Furthermore, if the evolutionary ethicist is saying that the genes which determine moral attitudes can, depending on environmental pressures, produce different, sometimes mutually incompatible, judgments – in other words, that our moral genes may in some environmental conditions produce the moral belief that *mp* and in other cases $\sim mp$ – the burden is on them to give an account of which environmental conditions produce which moral principles and why. It might be that at some point in the future, we will have answers to these questions. But until then, the realist need not accept that genes and environment are all there is to moral judgment. Moreover, there is still the question of the more in the moment disagreement where one is in two minds over *mp* and $\sim mp$. Such cases imply a stronger role for rational deliberation than the genetic-environmental account allows.

If my argument has so far been successful, I have shown that Street's evolutionary account of ethics fails to explain moral disagreement. I will now comment on why realists can offer a

better explanation. An argument from disagreement against some metaethical position claims that if that position is true, then we ought to expect moral agreement. That this expectation turns out to be false is what undermines the metaethical position in question. There is some plausibility in the claim that moral agreement is implied by the truth of moral realism in all its guises. If some version of non-naturalism is true, moral facts are non-ontological, abstract truths, akin to mathematical and logical truths and ascertainable by rational intuition. If naturalism is true, moral facts are natural facts, to be discovered empirically. In either event, the moral facts are capable of being recognised by utilising cognitive skills humans have been shown to possess. This accounts for why, in rationalist and empiricist domains, humans have to a large extent converged around the truth. The fact that this has not happened in ethics is what makes people question moral realism.

One issue with this picture is the distinction I have previously alluded to, between the realist's views on how moral judgments *are* formed and how they *ought to be* formed. Many realists concede that, although people ought to form moral beliefs in cognitively reliable ways, in practice this often does not happen. There is a wealth of literature on how our moral thinking is distorted by various biases and distortions,⁹⁴ as well as evidence that emotional factors often (but not always) override reason (Greene 2014, ch.4). In my view, we should reject any *monistic* normative genealogy which claims that normative beliefs are caused solely, or primarily, by one force, whether it be natural selection, culture, reason, or anything else. What we should claim, instead, is that normative beliefs are caused by a range of forces. I have already claimed that we should partially accept the evolutionary genealogy. Mackie's cultural moral genealogy is also correct to a point (no one can deny that culture plays a salient role in the transmission of normative values). We should cast the net wider than this. One of the ways in which normative beliefs differ from descriptive beliefs lies in their practical relevance, and as such, people might have non-moral reasons to accept or not to accept the truth of some normative principle. For instance, self-interested considerations have no apparent bearing on whether people come to accept mathematical propositions like $2+2=4$. But choosing to accept moral principles can be influenced by self-interest. For instance, suppose it is true that a system in which land is communally owned is *morally preferable* than one in which land is privately owned. The landowners might have a self-interested reason to not want the morally better option. Likewise, if I am well-off, I might have a self-interested reason to be against higher

⁹⁴ Haidt cites numerous biases: a bias in favour of the moral views of close ones; a tendency to be more sceptical of evidence which conflicts with one's own views than views with which one disagrees: Haidt (2001, p.821).

taxes on the rich. Thus, in general, there can be *more to lose* from accepting some moral principle than some descriptive principle. And this struggle between self-interest and morality might in fact be responsible for much moral disagreement. As further evidence of this, it is telling that those well-established descriptive facts that seem subject to the most disagreement tend to be the ones that have practical implications for the way we live our lives, such as disagreement about the truth of anthropogenic climate change.

In explaining normative disagreement, the realist should claim that our normative views are caused by many different factors pulling us in different directions, and therefore, there should be no reason, if we accept this pluralistic genealogy, to expect normative uniformity; that should be an expectation only if we accept a monistic genealogy. For instance, if person *a*'s belief that *mp* was caused primarily by self-interest, undue deference to authority or some other distortive force, whereas person *b*'s belief that $\sim mp$ was caused by reason or concern for others, it should not come as a shock that they ended up disagreeing. Furthermore, the realist will claim that some of these forces are truth-tracking and some are not. Unless we are ethical egoists, self-interested considerations will generally be distortive. As will, if we accept the premise of EDA, those normative beliefs influenced by natural selection. Given all these potentially off-track forces, we might wonder whether we could ever come to recognise normative truth. However, reason can prevail, even if it often does not. And though seldom the decisive factor, rational argumentation has played an important role in many of history's most morally appraised social movements.

Two examples will further emphasise this point that we can use reason to transcend distortive influences on our moral thinking. The first is one I have already discussed; namely, climate change, and the plausible thought that our inept response to it is partly because we evolved to be short-term thinkers. Nevertheless, some people take climate change very seriously, often bearing severe costs to themselves to mitigate its effects on future generations. Reason allows us to transcend our short-term bias by taking, to use Sidgwick's phrase, the "point of view of the universe", a perspective from which everyone's interests are worthy of equal consideration, irrespective of their temporal and spatial location (1907, p.382). A second example is charity. People are often moved by emotional factors when donating to charity, making them more likely to donate to identifiable recipients even if this is not the most effective distribution.⁹⁵ The evolutionary explanation is that we evolved in small groups in which our sympathies were

⁹⁵ See Singer (2009, Ch. 4) for a discussion of the biases which affect charitable giving.

guided by face-to-face contact with the recipients of our altruism from whom we could expect reciprocity; we did not evolve to act altruistically toward unidentifiable victims (Singer 2009, Ch.4). However, although we are susceptible to these biases, many people do make charitable donations on the basis of their effectiveness, without consideration of one's relation to the recipient or through an expectation of reward.⁹⁶ It follows that even if many of our moral beliefs are off-track due to epistemically distortive forces, there is evidence that human beings are capable of overcoming them.⁹⁷

4. Objection: Does This Argument Miss the Target?

Before continuing, I want to pause to consider a possible objection here. It might be argued that the argument of this chapter is something of a red herring, on the grounds that my interpretation of Street's position is misleading. My argument assumes that Street's account of evolution's relation to our normative judgments implies that evolution has shaped the content of our normative judgments in ways that were evolutionary advantageous. Hence the reason I claim that normative disagreement and the existence of a significant number of normative beliefs that lack an evolutionary explanation are a problem for Street's account. I acknowledged in chapter two that some evolutionary debunkers (content-neutral evolutionary debunkers) deny that evolution shaped the content of our normative beliefs. Thus, content-neutral evolutionary debunkers are not the target of the argument. The target of this chapter's argument are content-partial evolutionary debunkers, who believe that evolution shaped the content of our normative judgments and I have argued that Street falls within this camp. But perhaps this interpretation is mistaken, in which case, my argument misses its intended target. I will now spend some time defending my interpretation of Street based on the textual evidence.

At several points within her EDA against realism Street makes it clear she thinks that evolution shaped the content of our normative judgments. Street writes: "Evolutionary forces have played a tremendous role in shaping the content of human evaluative attitudes" (Street 2006, p.109). Moreover, it is not just purely attitudes as opposed to beliefs that evolution has had a hand in shaping. Street also goes on to write that "the forces of natural selection have had a tremendous influence on the content of human evaluative judgement" (ibid, p.113). These include, as Street mentions, "judgements about what is a reason for what, about what one should or ought to do, about what is good, valuable, or worthwhile, about what is morally right or wrong, and so on"

⁹⁶ For instance, effective altruists and blood and organ donors in nations in which this is not paid.

⁹⁷ For evidence showing how reason is often the main factor in moral judgments, see Greene (2014, Ch.4). For an account of how reason can transcend evolutionary biases, see Singer (2011).

(ibid, p.110). So it is not just that evolution provided us with tendencies to think normatively but that evolution has shaped the content of our normative judgments.

Moreover, Street seems to think that evolution shaped these judgments in a somewhat deterministic way, in so far as some judgments rather than others increased reproductive fitness. Street lists several normative beliefs of which she claims it is no accident we hold them, given that organisms who failed to hold them would struggle to survive in evolutionary terms. With this in mind, Street urges us to “note the potentially phenomenal costs and benefits, as measured in the Darwinian currency of reproductive success, of accepting some evaluative judgements rather than others” (ibid, p.114).

However, it is important to note that Street does not think that evolution *directly* shaped the content of our full-fledged normative beliefs. That is to say, Street does not think that fully-fledged normative beliefs are genetically heritable. What is genetically heritable, are the “proto versions” (ibid, 114) of those beliefs, which Street refers to as basic evaluative tendencies, which “may be understood very roughly as an unreflective, non-linguistic, motivational tendency to experience something as “called for” or “demanded” in itself, or to experience one thing as “calling for” or “counting in favour of” something else” (ibid, p.119). Thus, although Street is not saying evolution selected for normative beliefs, it is ultimately shaping the normative beliefs we came to hold because it shaped our initial normative attitudes. In the first instance, we, along with our close evolutionary cousins, directly evolved to hold basic evaluative tendencies to be emotionally and motivationally approving or disapproving of certain actions. Then later, as a result of later evolutionary add-ons such as advanced linguistic and rational capabilities these basic tendencies became fully-fledged normative beliefs.

However, Street claiming that evolution’s relation to our full-fledged normative beliefs is indirect in no way makes it immune from the arguments of this chapter. Indeed, I outlined in chapter two a very similar account of the way evolution shapes normative beliefs. My arguments work by undermining the claim that evolution shaped our normative beliefs in particular directions, irrespective of whether this shaping process was direct. In the case of my argument from non-adaptive normative beliefs, my claim was that the existence of normative beliefs which cannot be explained by evolution count against Street’s argument. This still applies, for according to Street, our fully-fledged normative judgments are by and large the product of our evolved basic evaluative tendencies. On Street’s account, we believe in things like kin partiality and the acts and omissions distinction ultimately because we directly evolved

an evaluative tendency in favour of them, without which, we would not have come to hold the fully-fledged normative belief.

But if our basic evaluative tendencies are those that were evolutionarily advantageous for us to hold, such as approval of kin partiality, the acts and omissions distinction etc., then this explains why people hold full-fledged beliefs like *one should give greater weight to the interests of kin* and *doing harm is worse than not preventing harm*. Consequently, this account cannot explain why some people do not hold those normative beliefs, for either, they did not evolve to have those original evaluative tendencies (which would mean their fully-fledged normative belief in *utilitarian impartiality* and *the rejection of the acts and omissions distinction* do not originate in evolution and consequently cannot be debunked by evolution) or they did evolve to have these basic evaluative tendencies, but that non-evolutionary influences such as reason or culture played some overriding role which resulted in their fully-fledged normative beliefs being at odds with their original evaluative tendencies (in which case, such fully-fledged normative beliefs would not even have been indirectly shaped by evolution and therefore, would not be debunkable by evolution). Ultimately, then, my arguments in this chapter target the claim that evolution, whether directly or indirectly, overwhelmingly shaped the content of our normative judgments in ways that were evolutionarily advantageous, and for reasons outlined, this includes Street's account of evolution's connection to our normative beliefs.

Conclusion

Street's content-partial EDA makes two key assertions. One is that if evolution shaped a normative judgment, it most likely did not track the truth. The other: that evolution overwhelmingly shaped our normative judgments. In this chapter, I did not discuss the first claim and for the sake of argument, I accepted it as being true. Street's Darwinian Dilemma, in which she asks the realist to choose between asserting and denying a relation between evolution and normative truth, relates to this claim, hence the reason I have not opted to take a position on that dilemma. I will address this first claim and take a position on the dilemma in the next chapter.

In this chapter, I have focused on Street's claim that evolution did overwhelmingly shape our normative judgments. I started by considering different interpretations of this claim. The most

extreme interpretation – one which Street does not seem to endorse – is that evolution shaped all normative beliefs. Even if this were true, it would not rule out the possibility that we could revise our distorted evolutionary normative beliefs and replace them with more epistemically viable ones. Alternatively, it might be that Street is claiming that most of our normative beliefs were shaped by evolution, but that some were not. If so, the debunking implications would only extend to the beliefs that evolution did shape. The outliers, however small a number, would be immune to Street’s argument. Which means that if a realist holds these beliefs to be true, their belief in realism is untouched by Street’s argument. To make this argument more concrete, I listed several normative beliefs that do not seem explicable by evolution, including most aspects of utilitarianism (though I will discuss the more complicated case of evolution’s relationship with the hedonistic theory of value in the next chapter). Finally, I presented a revised version of the argument from disagreement. Given that our ancestors were subject to virtually the same evolutionary pressures, we should expect – if evolution shaped our normative beliefs – normative uniformity. The proponent of the content-partial EDA cannot give a good account of why there is so much disagreement. Moreover, the realist can give a better explanation. As long as the realist does not attribute any one cause – whether that be reason, culture, empirical observation – to the formation of our normative beliefs, there is no reason to expect normative uniformity in the first place.

I just want to finish by noting the significance of this conclusion. If my argument so far is correct, then in some cases there will be no serious need for the realist to take a stance on the Darwinian Dilemma. If I am a realist whose normative beliefs cannot be explained by evolution, then it does not matter whether evolution is or is not related to normative truth. The need to take a stance on the Darwinian Dilemma arises only if my normative beliefs were shaped by evolution. In the next chapter, I consider the prospects of realists whose normative beliefs were shaped by evolution.

Chapter 4: Can Evolution Shape a Normative Judgment without Undermining It?

Introduction

In the last chapter, I contested variations of the claim that evolution shaped most of our normative beliefs. This provided scope to reject the conclusion that all versions of realism are undermined by EDA, since realists whose first order normative beliefs cannot be explained by evolution are immune from EDA. I did not discuss Street's assumption that if evolution did shape some normative belief, it is unlikely to be in the direction of (stance-independent) truth. I will now consider whether this claim is true. I start by considering what it means for evolution to shape a normative judgment, for although such a claim seems straightforward enough, there is an important conceptual distinction between *merely causing* and *shaping* a belief and this distinction has relevance for Street's argument. I will then argue that for the most part, Street is on the right track in asserting that an evolutionary explanation of some belief tends to undermine it. However, I identify a category of normative beliefs which serve as counterexamples to this general trend. These are normative beliefs about our mental states, such as the hedonistic belief that pain is bad. In concerning our mental states, of which we have direct acquaintance, these beliefs are more epistemically secure than other 'evolved' normative beliefs, which are often more 'layered' in that they involve an intermediary emotional mental state. In defending this view, I will go some way in defending a hedonistic theory of disvalue from EDA, which in conjunction with my defence of utilitarianism in the last chapter, will mean that I have defended *Classical Utilitarianism* against Street's EDA.

1. Evolutionary Relations

1.1. Searching for the Problematic Relation

Street phrases the "Darwinian Dilemma" as follows:

The basic problem for realism is that it needs to take a position on what relation there is, if any, between the selective forces that have influenced the content of our evaluative judgements, on the one hand, and the independent evaluative truths that realism posits, on the other. Realists have two options: they may either assert or deny a relation (Street 2006, p.121).

Presenting this as a binary choice masks what is a thorny terrain in which there are several different paths open to the realist. In the following discussion, I will try to flesh out these different options before choosing the one I think offers the realist most hope in responding to Street's Darwinian dilemma. To begin, I will clear up a potential ambiguity in the above quotation: it is unclear exactly what question Street has in mind when it comes to the kind of relation the realist is being asked to consider. Street could be asking one of two questions:

- (1) Is evolution related to normative truths?
- (2) Is evolution related to our ability to know normative truths?

These two questions should not be conflated. Evolution can be causally related to some truth. For instance, the fact that hedgehogs are spiky mammals is true only because of the evolutionary history of hedgehogs. However, some truths are unrelated to evolution, such as the truths of physics. A way of establishing whether some truth t is related to evolution e is to consider whether t would obtain in the possible worlds in which e did not. Consider the truths of physics again. Even if the process of evolution never occurred, Newton's theory of Gravity would still be true (though Newton would never have been around to discover it). However, this is not true of the proposition that hedgehogs are spiky. For if the process of evolution never began, there would be no such thing as a hedgehog. Then there are cases in which e is not related to t but to our knowledge that t . One example of this is evolution's relation to our knowledge of mathematics. $2+3=5$ is true in all possible worlds; including worlds in which life never evolved. However, in the actual world, the one in which human beings can engage in mathematics, it is highly likely that this ability came about through natural selection; thus, evolution is unrelated to the truth of mathematical claims but is related to our knowledge of them.

Answering both (1) and (2) affirmatively in relation to some truth entails no inconsistency (the truth that hedgehogs are spiky mammals and our knowledge of that truth are related to evolution). But one can also consistently affirm one of the two relations whilst denying the other; in the moral context, this might be the view held by some contemporary religious people who see moral truths as divine commandments (and thus independent of evolution) yet accept that human beings and their cognitive functions are products of natural selection. Of these two questions, Street is interested in (2). She is not asking whether there is any relation between evolutionary pressures and the normative truths themselves. Rather, she assumes, for the sake of argument, that there are such truths. Her objection is that given the causal influence of

evolution on our normative judgments – a causal influence governed by the maximisation of genetic reproduction – evolution shaped our normative judgments in “evaluative directions that have nothing whatsoever to do with evaluative truth” (Street 2006, p.121). Under such conditions, it is unlikely that evolution placed us in cognitive reach of normative facts. It follows, then, that in asking the realist to assert or deny an evolutionary relation, the realist is being asked to decide whether the forces of evolution are related to the supposed ability of human beings to know, discover, or apprehend moral truths, not to the truths themselves. For now, I will focus on what it might mean to affirm such a relation and on Street’s reasons for objecting to such an affirmation. A candidate assertion of the aforementioned relation, which Street considers a problematic relation, is the following:

The Problematic Relation: Human knowledge of stance-independent normative facts is causally related to the forces of natural selection.

Though this assertion appears to be a simple and unambiguous statement of the kind of relation Street criticises, there is an underlying vagueness. Street wants to claim that there is a genuine choice for realists when it comes to affirming or denying an evolutionary relation. But framed like this, there is no real choice, since no realist who accepts the theory of natural selection would deny that this kind of relation obtains between the forces of evolution and our ability to recognise moral facts. This is because if it were not for evolution, there would be no humans, and in possible worlds in which humans do not exist, no humans possess normative knowledge. Therefore, all realists who accept the theory of evolution must accept that normative knowledge and the forces of natural selection bear some causal connection. Street, therefore, must have a more specific kind of causal relation in mind.

1.2. The Tracking Account of the Relation Between Evolution and Normative Truth

We get insight into the relation Street has in mind by considering the realist accounts she objects to. For instance, Street spends some time criticising Derek Parfit’s account of evolution’s connection to ethics. Parfit,⁹⁸ she claims, is one of a number of realists who assert what she calls a “tracking relation” between evolution and normative knowledge; on this account, evolutionary forces “tracked the truth” in the normative domain, presumably because normative knowledge was fitness enhancing (Street 2006, pp.125-127). Of the realist who takes

⁹⁸ Street’s discussion of Parfit’s view here is not based on his published work but on correspondence between the two. Street includes a quotation from Parfit (pp.125-126) which captures his tracking view: “just as cheetahs were selected for their speed, and giraffes for their long necks, the particular feature for which we were selected was our ability to respond to reasons and to rational requirements.”

this line, Street goes on to write that “surely, he or she might say, it is advantageous to recognize evaluative truths; surely it promotes one’s survival (and that of one’s offspring) to be able to grasp what one has reason to do, believe, and feel” (ibid, p.125). This is the kind of relation Street considers unacceptable. First, she finds it hard to believe that the abstract normative truths in which non-naturalists believe could have any such survival advantage, writing that “a creature obviously can’t run into such truths or fall over them or be eaten by them. In what way then would it have promoted the reproductive success of our ancestors to grasp them?” (ibid, pp.130-131). She also thinks that this position holds less explanatory power than her view that we hold the normative beliefs we do solely because such beliefs were advantageous, not because they are true (the *Adaptive Link Account*). Thus, when Street asks the realist to assert or deny a relation between normative facts and evolution, she has in mind a relation whereby evolutionary forces tracked moral knowledge because of the reproductive advantage it conferred on our ancestors:

The Problematic Relation*: Evolution selected for moral knowledge because such knowledge was adaptive

There are good reasons for being sceptical about this kind of relation between evolution and normative knowledge for the reasons given in chapter two. It is not obvious why it would be in our evolutionary interest to apprehend an abstract normative fact, assuming there are such facts. I will, as indicated, argue shortly that hedonistic normative claims are possible exceptions to this rule, but with that caveat noted, Street’s claim that asserting a normative truth tracking relation, at least of the kind above, seems to be a questionable move for the realist to make.

However, it is worth noting that the relation posited above is of a very direct kind. It asserts that normative knowledge as such, is adaptive, hence the reason evolution tracks it. But there is an option for the realist to assert an indirect tracking relation, which does not suppose, unscientifically, that normative knowledge is fitness enhancing. According to this view, our ability to recognise normative truths is a by-product of some other ability that is fitness enhancing.⁹⁹ The most widely endorsed by-product account is *rationalistic*. It claims that evolution selected for an ability to engage in various forms of reasoning, whether that be deductive, mathematical, inductive, because these abilities enabled us to form true beliefs that were fitness enhancing (how many predators are nearby, the fact that tigers eat people etc). According to Singer and de Lazari-Radek (2014, 179-185), the massive investment evolution

⁹⁹ For by-product accounts, see Parfit (2011b, pp.511-542) and Singer and de Lazari-Radek (2014, pp.179-185).

put into this ability was all or nothing. It was not that evolution selected piecemeal for us to know the pieces of knowledge that would be fitness enhancing. More likely, in their view, the initial investment was made to enable us to form true beliefs that were fitness enhancing, but having acquired the ability, we were able to then use it to form true beliefs about things that do not provide us with an evolutionary advantage, e.g., quantum physics, philosophy and most importantly for our purposes, normativity. The realist who endorses this account is not suggesting knowing about normative facts is fitness enhancing, only that this epistemic ability was a by-product of some other ability that was.

One response at this point might be to question the suggested trajectory from a basic ability to reason to an ability to recognise normative truth (see Street 2006, pp.142-144). Evolutionary psychologists claim that our evolved abilities are flexible and can be used for purposes other than the ones they originally evolved for. No one would claim, for example, that we evolved to play tennis, but we can easily explain this ability in terms of physical adaptations that were fitness enhancing for other reasons, such as to hunt prey and escape predation. Likewise, in the case of knowledge. Although evolution did not directly select for an ability to do advanced mathematics or quantum physics, we can explain those abilities as an extension of a more basic ability to perform mathematical calculations and engage in inductive reasoning. But, it might be argued, the realist should not claim that normative reasoning is an extension of any of these other forms of reasoning because it is different *in kind*.

The realist has two possible responses to this claim. First, they might question whether our knowledge of, for instance, quantum physics really is just an extension or application of the basic ability to make good inductive inferences about our surroundings. For one thing, some occurrences at the quantum level seem to conflict with our folk epistemology, in which case this ability might need to be explained by the plasticity of our reasoning capabilities and how they can be applied to a wide array of domains, including ones for which they were never 'intended'. Secondly, I do not think that the forms of reasoning associated with normativity are different in kind. The view I sketched out in chapter 1 proposes that normative facts are abstract and hence, amenable to the kind of deliberations philosophers are engaged with. It is an empirical fact that philosophy is an intellectual practice performed in most, if not all, cultures. It seems very plausible that the ability to engage in philosophy and acquire philosophical knowledge originated in the innate reasoning capabilities we evolved to possess, so why not an ability to recognise normative truth?

I will now consider two responses to this line of thought. First, the anti-realist could double down on their anti-realism. They might grant for the sake of argument that normative reasoning is a branch of philosophy but deny that philosophy is in the business of discovering truths. Unfortunately, I have not got the space to defend realism about philosophy (the belief in philosophical facts), and as such, I acknowledge that my line of argument could be derailed by the truth of some kind of philosophical *quietism*. However, the problem with the anti-realist making this argument is that if it works, they become a victim of their own success.¹⁰⁰ Metaethics is a branch of philosophy. The claim that *normative realism is false* is a metaethical – and consequently philosophical – statement. Which means that if philosophical anti-realism is true then the evolutionary debunker’s conclusion cannot be true.¹⁰¹ A second anti-realist response is to deny that normative reasoning is a species of philosophy. But as I mentioned in the first chapter, it does share a lot of the hall-marks of philosophical reasoning. For one thing, neither are in an obvious way engaged in inductive reasoning. The method for establishing or undermining propositions in both areas is basically the same, by checking for consistency, appealing to our intuitions about thought experiments and engaging in rational argument. I say more about the similarities between philosophy and normative thinking in the next chapter.

Finally, leave aside for a moment the question whether normative claims can be true. Street and Joyce are making metaethical assertions about the nature of normativity. Their assertions, if true, would be philosophical truths. Though they do not seem to have expressed a view on this point, it would be interesting to know what explanation they would give for our ability to recognise philosophical truths. It is unclear what survival advantage being good at philosophy would provide. One could respond that being able to reason validly might have been advantageous and logic is usually considered a branch of philosophy. True, but there is also something different about logic, both in the certitude of its propositions and the way we accept them on self-evidence rather than argumentation. If anything, our recognition of logical truths seems more akin to apprehension of basic mathematics. It is not immediately clear how being good at logical deduction would enable us to answer - though of course it plays a role in answering - questions like ‘Do we have free will?’, ‘What is knowledge?’ and ‘are there normative facts?’ I fail to see how being able to know the answers to these questions would provide a selective advantage. And indeed, analogous to the case of ethics, evolution might

¹⁰⁰ This point is made by Shafer-Landau (2012, p.25).

¹⁰¹ This is basically a variation on the argument that scepticism is self-defeating. According to a thoroughgoing sceptic, *none of our beliefs are justified*. But if *none of our beliefs are justified*, then the belief *that none of our beliefs are justified*, is not justified. Which means we have no reason to accept the sceptic’s claims.

sway our views in the wrong direction (e.g., irrespective of whether we have free will, it might be advantageous to believe so, but surely that does not mean that attempts to answer this question are futile). We can now present the evolutionary anti-realist with the following dilemma. Either:

- a) Evolution selected for an ability to recognise philosophical truths

Or,

- b) Evolution did not select for an ability to recognise philosophical truths

By taking (a) the evolutionary anti-realist could explain how they came to recognise the metaethical truth that evolution undermines normative realism. The problem is that by Street's own lights, (a) is unscientific. To loosely paraphrase Street's earlier remarks, we cannot fall over or be eaten by abstract philosophical (or metaethical) truths, so why would it make evolutionary sense to grasp them? If Street chooses (b) then we can ask by what means she was able to recognise the philosophical truth of anti-realism if evolution did not provide us with an ability to recognise such truths. Street would need to say that some factor other than evolution was responsible. But in claiming this, it seems arbitrary to claim that such a response is not available to the normative realist in explaining how they came to recognise normative truth. A final possibility. Let's suppose that there are philosophical truths, including metaethical truths. Let us also suppose that evolution shaped our philosophical and metaethical beliefs. Given that there is no survival advantage to be derived from recognising such truths, if evolution did shape our philosophical beliefs, it most likely did so in ways that do not track philosophical truth, meaning that only a very unlikely coincidence could have led to us recognising them (Shafer-Landau 2012, p.25).

1.3. Causing, Shaping, and the Darwinian Dilemma Revisited

To affirm or deny a relation between evolution and normative knowledge, that is the question we started with. The position I am taking is that evolution is related to our capacity for normative knowledge, but usually in an indirect way, by providing us with an ability to engage in complex forms of reasoning which we were then able to use to form true beliefs about things that provided us with no adaptive advantage. The fact that the relation I posit is indirect is important. Street claims that evolution overwhelmingly shaped our normative judgments. In claiming that evolution is indirectly related to our ability to acquire normative knowledge, am

I claiming that our true normative beliefs were shaped by evolution? Not necessarily. x can be causally related to y without shaping the nature of y . Consider the following two scenarios:

- (1) It's my birthday. As a gift, my good friend has given me a new set of watercolours. I use them to paint my favourite Lake District Mountain.
- (2) A violent and extraordinarily vain criminal breaks into my house, gives me a set of watercolours and demands, at gun point, that I paint *him*.

In the first case, my friend's actions were causally linked to my painting in that, had she not given me those watercolours, I would not have painted the mountain. But I do not think we should go as far as to claim that my friend's action shaped the content of my painting. This was shaped primarily by my aesthetic appreciation of my favourite mountain. In the second case, not only were the criminal's actions causally related to my portrait of the criminal, but they shaped the nature of my painting: had he not demanded at gun point that I paint him, I would have had no inclination to do so. Consider a real example: Wordsworth's *The Prelude* (1850). Did evolution shape the nature of this poem? I think not. Although evolution was causally related to the fact Wordsworth wrote *The Prelude* – both because it provided humans with the linguistic capabilities to write poetry and it gave us Wordsworth – the poem itself was shaped by factors such as Wordsworth's autobiographical reflections and artistic sensibilities.

We can say something similar about normative knowledge. I think that Sidgwick's axioms, some of which I discussed in the last chapter, are true. Evolution is causally related to my knowledge of them, because I would not exist (nor would Sidgwick) and be able to comprehend these axioms had it not been for evolution. But I do not think that evolution *shaped* my beliefs in the direction of Sidgwick's axioms. Perhaps evolution sometimes does this. I suspect the belief that we should put family members' interests above strangers' interests probably was both caused and shaped by evolution. But when it comes to normative beliefs like those discussed in the previous chapter, which do not lend themselves to an evolutionary explanation, we should affirm only a causal but not a shaping relation. Does this make any difference as to whether this relation serves to undermine the belief in question? I think it does. Let's suppose, along with Street, that evolution is indifferent toward us being able to recognise normative truth, and as such, most likely shaped normative beliefs in directions that do not correlate with truth. We can easily see how such beliefs might end up off-track. But where evolution is merely a causal influence, not a shaping one, it did not shape our beliefs in any direction, let alone the right or wrong direction.

Before proceeding, I want to make it clear that I am not denying that evolution shaped some of our normative beliefs. Just as evolution gave us a propensity to accept certain basic empirical and mathematical claims because they helped our ancestors survive in their Stone Age environment, it also provided us with certain normative dispositions. Our original ‘folk’ normative beliefs served an important practical purpose, but not necessarily one pertaining to truth. But it does not follow that we cannot make up the gaps and revise our mistaken beliefs. The fact that quantum physics often conflicts with our folk understanding of science is not an argument against quantum physics. I largely agree with Street that evolution has massively shaped our normative thinking. Much of what it gave us might have been useful, though ultimately off target. On the other hand, within a sea of distortion it might have left us with some grains of truth. As a species whose reasoning powers are not wholly bound by evolutionary designs, there is no reason why we cannot search and hopefully find these grains of truth within the framework of our evolved normative beliefs, or indeed, look for them elsewhere.

In the last chapter, I argued that evolution did not shape all our normative beliefs and therefore, *some* normative beliefs cannot be undermined by evolution. So far in this chapter I have noted an ambiguity in the claim that evolution caused us to hold some normative belief. If the form of causation was indirect, in that evolution provided us with the cognitive abilities which enabled us to form true normative beliefs, then we presumably would want to say that such beliefs are not really shaped by evolution, and that, so long as the more proximate cause of those beliefs is truth-tracking, evolution does not undermine that belief. I cited the proximate cause of utilitarian intuitions to be our ability to reason (a belief for which there is empirical evidence).¹⁰² Our reasoning abilities, though fallible, are known to track the truth. Which means that normative beliefs whose proximate cause is our ability to reason are not obviously off-track.

This still leaves the question of whether beliefs that are more directly shaped by evolution, such as kin partiality, are undermined by that explanation. I gave some grounds in the second chapter for why we might answer in the affirmative. Forming a belief by unreliable means is unlikely to lead to truth, hence our scepticism of the person whose time telling was informed by a faulty clock. In the case of evolution, I conceded that the evolutionary debunker is on the right lines in claiming that the process of evolution is unlikely to directly select for normative knowledge.

¹⁰² I outline some of this evidence in ch.5 when I discuss Joshua Greene’s brain scans of subjects as they carry out moral judgments.

The reason being that there is no obvious explanation for how such knowledge would help an organism survive.

Whether evolution undermines the normative beliefs which it *directly shapes* will be the focus of the remainder of this chapter. I think for the most part, if evolution shapes a normative belief, it does tend to undermine its credibility. However, I deny that this statement holds universally and I will seek to identify a counterexample. In the last chapter I claimed that the fundamentals of utilitarianism (*universal benevolence, value maximisation, rejection of the acts and omissions distinction*) cannot easily be explained by evolution. But in so doing, I did not fully defend the utilitarian view I endorse for reasons outlined by Guy Kahane when he tells us that “[...] utilitarianism is empty of content unless supplemented by an account of well-being. But many of our evaluative beliefs about well-being, including the beliefs that pleasure is good and pain is bad, are some of the most obvious candidates for evolutionary debunking” (Kahane 2011, p.120). The utilitarian account I am defending is hedonistic. And whilst I think Kahane is right that a hedonistic account of value is a *prima facie* candidate for evolutionary debunking, a nuanced argument can be given in defence of it. I will now attempt to demonstrate this.

2. Evolution and Hedonistic Disvalue

Before proceeding with this argument, a brief digression into the philosophy of pain¹⁰³ is necessary. The reasons are as follows. First, according to most definitions of hedonism, the only thing of intrinsic disvalue is pain. However, I do not consider this claim necessary for one to qualify as a hedonist. Moreover, it does not happen to be the hedonistic account of disvalue which I accept. Perhaps of more importance, I do not think all hedonistic accounts of disvalue are immune to EDA; or at least, some are more immune than others. And it is for these reasons that some time needs to be taken in outlining my hedonistic account of disvalue. A final point before proceeding. Hedonism is the monistic view that *only* certain mental states (usually taken as pleasure) are good and that only mental states (usually pain) are bad. A non-hedonist might accept that *pain is bad* and *pleasure is good*, but hold the pluralistic view that other things are intrinsically good (or bad) for us too (such as knowledge, friendship, aesthetic appreciation etc.). A positive argument for hedonism would need to show that hedonic mental states alone are (dis)valuable. That is not my intention here. My aim is to defend hedonism against EDA. And EDA typically focus on showing that judgments such as *pain is good* and *pleasure is bad*

¹⁰³ The *philosophy of pain* covers a range of topics relating to pain. Such as: Is pain intrinsically or instrumentally bad? What makes pain bad? Is pain necessarily bad? Is pain objectively bad?

can be explained by evolution, rather than the hedonist's claim that nothing other than these states is of intrinsic value. Consequently, my focus will be on showing that the mental states hedonists ascribe (dis)value are not debunkable by evolution, rather than the claim that nothing other than these states possess intrinsic value.

2.1. Realism and Pain's Badness

Realists assert that some normative propositions are true. If I am a realist who holds that *pain is bad*, then I take *pain is bad* to be an assertion of fact rather than an expression of some non-cognitive attitude toward pain. A realist about pain's badness will claim that this fact is stance independent, meaning it is true regardless of what anyone thinks about pain's badness. Next there is the question of whether pain is intrinsically bad. Some theories of wellbeing claim that pain's badness is instrumental. A notable example being the desire-fulfilment theory, which claims that pain is bad only if one desires not to be in pain. This also means that the desire fulfilment theorist denies that pain is *necessarily* bad, for in the absence of a desire to be pain free, it is presumably neutral, and in the presence of a desire to be in pain, being pain-free is presumably of negative value. Hedonism has historically been thought of as a normative theory which takes pain's badness to be intrinsic. This perspective is supported by certain classical and contemporary expositions of the view.¹⁰⁴ However, although some versions of hedonism make this claim, one need not claim that pain is intrinsically, or even universally bad, in order to qualify as a hedonist. To further substantiate these remarks, we will need to consider the question of *why pain is bad*.

2.2. Why is Pain Bad?

One highly influential and intuitive theory of pain's badness is the *Sensation Theory*.¹⁰⁵

The Sensation Theory: Pain's badness lies in the qualitative feeling of pain.

The sensation theory of pain is in opposition to functionalism. Functionalism emphasises the role pain plays in indicating the presence of bodily damage.¹⁰⁶ However, the objectively observable features of pain – the firing of C-fibres, wincing, nerve damage, and so on – cannot adequately explain pain's badness, because it only tells one side of the story. Pain is not a

¹⁰⁴ Mill writes: "Those who know anything about [utilitarianism] are aware that every writer, from Epicurus to Bentham, who maintained the theory of utility, meant by it, not something to be contradistinguished from pleasure, but pleasure itself, together with the exemption from pain [...]" (Mill 1863/2015, p.120).

¹⁰⁵ For defences of the sensation theory, see Rachels (2000) and Goldstein (1989). Rachels uses the term "Intrinsic Nature" for what I have described as the sensation theory.

¹⁰⁶ See Hall (1989).

wholly objective phenomenon.¹⁰⁷ Indispensable to pain is having a certain kind of subjective experience. Moreover, there is reason to think the badness of pain is more contingent on its subjective features. Would there really be anything bad about the firing of C-fibres if it did not produce any negative experience? Probably not. Would there be anything bad about the feeling of pain if it was not accompanied by its physical concomitants? I think so. The awfulness of an intense migraine would not be mitigated if we were to discover it had no physical symptoms. A final point in favour of the sensation theory of pain is that it seems to fit neatly with our common-sense intuitions: pain is bad *because it hurts!*

However, the problem with this view is that pain sometimes does not hurt. There are people – pain asymbolics – who can experience the sensation of pain, but do not seem to mind that experience (see Bain 2014, pp.305-320). Given the indifference asymbolics exhibit toward their pain, it is not obviously bad for them. Moreover, as Gwen Bradford has argued, many of us seem not to mind or even enjoy some pains; to use her examples, “eating very spicy food, getting a deep tissue massage, jumping in a freezing lake, sitting in a very hot sauna, or eating something with a strong bitter or sour flavour” (Bradford 2020, p.239). Bradford calls these “Hurts So Good” experiences (ibid). It is worth pausing here. Some experiences might be composite mental states containing a slight surplus of pleasure over pain – for instance, bathing in very hot water. Of such an experience, we might not want to say that we enjoy the pain associated with bathing in hot water, but that the pain is mildly outweighed by the pleasure, which is why we might consider this experience good. Other times the pain seems to be part of the enjoyment. I enjoy the moderate burning sensation in my thighs as I hike up mountains or ride my bicycle, people enjoy the fear induced by a horror-movies, or the sadness invoked by a tragedy. In these cases, the pain-like sensation seems to be an indispensable part of the enjoyment we derive from the experience.

The sensation theorist could persist in their objection, claiming that despite appearances, these are still mixed mental states containing both pleasure and pain, and that our enjoyment of them lies in there being a surplus of pleasure over pain, however marginal. There is some truth in this: the aforementioned experiences do involve a mix of pleasure and pain. I would object that a sensation resembling pain can go from being bad to good (or visa versa) without any change

¹⁰⁷ As Nagel argued, when it comes to consciousness, purely objective explanations are insufficient because subjectivity is an indispensable aspect of consciousness. We humans might know everything about the physiology of a bat’s brain, but that cannot tell us what it is *like to be a bat* because for that we would need to experience bat consciousness. Hence the need for a *Dual Aspect Theory* which gives appropriate weight to both the subjective and objective aspects of mind. See Nagel (1986, Ch.2) and (1974).

in its *qualitative phenomenology*, apart from its level of intensity. As I cycle along the flat surface, I feel but a mild burning in my legs, which I not only do not mind, but seem to enjoy. As the incline gets steeper, the burning increases, it goes from being mildly enjoyable, to being about neutral, to being a mild nuisance, to being – eventually – excruciating. Even the case of bathing in hot water seems to show this. I like the feeling of the hot water to some degree, but beyond a certain threshold my response – depending on the quickness of the change – gradually turns from positive to negative. Is there any real difference in the qualitative nature of this sensation, rather than in its intensity? If so, when does the qualitative change occur? The answer cannot just be *when I start to dislike it* for that would imply that it is not the mere sensation of pain that matters, but our response to it. Moreover, to draw a rigid line denoting when an experience of this kind becomes painful seems arbitrary. I'd suggest that the sensation does not change in any fundamentally qualitative sense – it is, qualitatively, the kind of sensation we associate with pain. A little bit of it does not bother us, in fact, we might even like it, but beyond a certain point, most of us dislike it, with notable exceptions such as pain asymbolics, who are indifferent toward it. A benefit of holding this view is that we need not draw some arbitrary line nor claim that the mere sensation of pain is bad, which it does not seem to be in the case of people like pain asymbolics.

Another response might be to deny that something can count as pain if it does not hurt the subject who experiences it. But consider again the case of pain asymbolics. Most pain asymbolics previously felt an aversion to pain. Thus, they are well acquainted with what the sensation feels like. It would be unduly sceptical to claim that they are mistaken about having a pain-like mental state given their familiarity with that mental state on previous occasions. Perhaps the sceptical argument is not bringing into question the asymbolics' knowledge of their mental states but is simply making a conceptual point about the correct analysis of pain: a necessary condition for a sensation's being painful is that it hurts. But what does it mean for a pain to hurt? Presumably it means we dislike or are averse to pain. In which case, it is not just pain purely conceived as a raw sensation that is bad, but our response to it that determines its badness. This thought has led some philosophers to adopt a *Dislike Theory of Pain*.

The Dislike Theory: Pain is bad when and because it is disliked.

As Derek Parfit puts it: "When we are in pain, what is bad is not our sensation, but our conscious state of having a sensation we dislike" (Parfit 2011a, p.54). The dislike theory can more easily make sense of asymbolics and those pains which seem not to diminish our

wellbeing. It does this by denying that pain is always bad, by making its badness contingent on being disliked. However, phrased as it is above, the dislike theory seems to conflict with both realism and hedonism. It conflicts with realism because it makes pain's badness contingent on the stance we take toward pain. It conflicts with hedonism because it seems to violate the *Experience Requirement* which posits that for something to be of disvalue, it must have some bearing on our experiences. Perhaps these objections are mistaken. To dislike something involves a particular kind of mental state and therefore a type of experience, so perhaps it does not violate the experience requirement. There is an issue about the word *dislike* which is ambiguous and seems to denote numerous mental states that are qualitatively distinct. For instance, I can dislike the taste of kimchi and I can dislike some governmental policy, but these dislikes seem different *in kind*. If I claim that I dislike kimchi, I mean to describe a certain form of qualia whose experience I dislike. But when I say I dislike some governmental policy, I am not claiming to dislike the experience I have when I learn or think about this government policy (though I may dislike this qualia also) but the actual policy itself, which is external to my mental states. Furthermore, the dislike theory as formulated above seems to allow for my disliking pain for reasons which have little to do with how it bears upon my experiences. For example, I might oddly dislike all sensations beginning with the letter "P" and this may form the basis of my dislike of Pain. Surely, if a dislike theory of pain is remotely hedonistic, it must make our dislike of pain contingent on the way it feels. A thought which might lead us to conclude that hedonists should be sensation theorists about pain.

Some philosophers have developed the dislike theory in ways that make it more compatible with hedonism. Derek Parfit draws a distinction between hedonic and meta-hedonic disliking (Parfit 2011a, pp.52-56). A hedonic-disliking entails disliking some experience solely because of the way it feels at the moment we feel it. This kind of disliking is instinctive and non-cognitive, an example being the brief few moments of awfulness which may accompany stepping on a mousetrap. We can also dislike things in a more cognitive manner, involving mental states such as *judgments*, *beliefs* and *desires*. These are what Parfit calls meta-hedonic dislikings. Parfit notes that, "When we are having some sensation that we intensely like or dislike, most of us also strongly want to be, or not to be, in this conscious state" (Ibid, p.54). But there is a conceptual difference here. The hedonic disliking is the complex mental state of disliking some experience I am currently having. The meta-hedonic disliking involves taking an attitude toward that experience, by for example desiring for an end to being in the conscious state of disliking my pain, or being motivated to end this experience. But note that the hedonic

dislike is an *experience* whereas the meta-hedonic dislike is an *attitude, desire or motivation* pertaining to that experience.

Parfit's distinction does important work in capturing the sense in which we usually dislike pain. There is something *in the moment* about disliking pain compared with other dislikes. For instance, if I experienced intense pain a week ago, assuming that pain has now ceased, I no longer dislike that pain (though I might have some bad memories about the period of time when I was in pain). And it is important to distinguish between things that I dislike because of how they feel and things that I dislike for other reasons. A hedonist might argue that pain's badness involves having a hedonic, as opposed to meta-hedonic disliking of it; the latter pertaining more to attitudes, judgments, desires (and thus more inclined to response-dependent views) the former to experiences (and thus more in line with hedonism). Perhaps though, given the vagueness of the word dislike, we would be better to find a more precise alternative. Kahane, sensitive to this ambiguity, prefers to use the word *aversion* to describe our negative response to pain. According to his *Aversion Theory*, pain's badness lies in the "affective state of finding that sensation aversive" (Kahane 2017, p.213).

I agree with Kahane that the word aversion does a better job of capturing the experiential nature of our negative response to pain. But I think that, insofar as there is any difference between his aversion theory and Parfit's hedonic dislike theory, it is basically semantic. Both seem to agree about what makes pain bad and about what kind of experience has disvalue, the differences seem to be about how best to describe that experience. As I am being electrocuted, the first thing that captures my consciousness is the sheer awfulness of the experience, whether we want to call it an intense disliking, an aversion, a negative experience, undesirable consciousness, as with many of our experiences, words often fail us. I think it would be too strong to call this intense aversion a *stance*, if by that you mean any kind of cognitive attitude, judgment, opinion, evaluation or something similar. Such cognitive states may be caused by my aversion, such as a desire or motivation for pain to stop. But I can conceive of aversions without them. Brief and intense pains like stepping on a mousetrap do not seemingly allow space for the formation of a desire or a motivation for the pain to stop. And pains can still be bad without any accompanying motivation or desire: if some being, such as a new-born baby or a certain species of animal did not have the cognitive awareness required for *motivation* or *desire* it would still be a terrible thing if they were to experience excruciating pain.

The Aversion Theory of Pain: The experience of being in a state of aversion toward one's pain is bad (during those moments when one is in pain).

Can the aversion theory be realist? Can the aversion theory be hedonistic? Let's start with the second question. Though one can accept the aversion theory of pain without being a hedonist – e.g., a pluralist who thinks both mental states and non-mental states can have disvalue – the aversion theory is compatible with hedonism. Hedonists claim that only certain experiences can be good or bad for people (the experience requirement). Often this is cashed out in terms of pain and pleasure, but it need not be. Sidgwick provides to my mind one of the most precise definitions of the hedonistic account of *the good* as that “feeling which the sentient individual at the time of feeling it implicitly or explicitly apprehends to be desirable; - desirable, that is, when considered merely as feeling, and not in respect of its objective conditions or consequences” (Sidgwick 1907, p.131). Note that the word desirable does not mean *actually desired*, but a feeling whose quality gives us *reason to desire it*. Following Sidgwick, we might claim that experiences of *disvalue* are those which we apprehend as being *undesirable* at the time at which we experience them. This raises the question: What experiences are undesirable? The answer, for reasons given, cannot be the *mere sensation* of pain, because that sensation is sometimes not apprehended as being undesirable. The experience of being in a *state of aversion toward one's pain*, however, is apprehended as undesirable for all those who experience it. And unless we are to claim that *being in a state of aversion to one's pain* does not qualify as an experience (which is a very odd claim indeed) I see no reason as to why one who proclaims that this experience is the only thing which serves to diminish wellbeing should not qualify as a hedonist given that they pass the experience requirement with flying colours.

What about the aversion theory's compatibility with realism? Realists claim that *the bad* is stance-independent. A realist who accepts the aversion theory can argue:

The only thing which has disvalue is a certain kind of experience. That experience is being in a state of aversion toward one's pain in the moments when one is having that experience. Even if someone were to form the judgment that *there is nothing bad about the experience of being in a state of aversion toward one's pain* or had no desire to avoid this experience, it would still be bad for them. In which case, the badness of being in a state of aversion toward pain is stance-independent: having this experience would be bad even if everyone in the world denied it or had no desire or motivation to avoid this experience.

The subjectivist might respond:

In conceding that pain's badness is contingent on an aversion or dislike, you are conceding to subjectivism. Because to dislike something is to take a stance against it.

However, this objection misses the point. It is true that the aversion theory makes pain's badness stance-dependent. But the above hedonist is not claiming that pain is *the bad*. Rather, she is claiming that *the bad* is being in a state of aversion toward one's pain at the time in which it is felt. So, whilst I accept that pain's badness is stance-dependent, I do not accept that *the bad* – which for me is being in a state of aversion toward pain - is stance-dependent. This most likely will not satisfy the subjectivist. Even though this hedonist is claiming that *the bad* is a certain kind of experience, the experience in question has a stance - being in a state of disliking - baked into it. But again, I wonder whether the issue here is the ambiguous nature of the word *dislike*. My preferred word *aversion* seems better at capturing the raw negative response to pain which I have in mind. Would we really want to say that a rat suffering from electrocution is taking a stance or attitude toward their plight? Not if that word is to mean a complicated mental state involving beliefs or desires, as opposed to a raw aversion. I accept that this may not be an all or nothing distinction. Subjectivity and objectivity can be cashed out in different ways. I noted in chapter one that on the "mind-independence" account of realism, hedonism would fail to make the grade because it makes wellbeing contingent on mental experiences. On the "stance-independent" formulation which I favour, I accept that my version of hedonism is *somewhat less* objectivist than 'ordinary hedonism,' in that it makes badness more contingent on our responses to pain. At the same time, it is clearly more objectivist than views like subjectivism and relativism which make wellbeing more of a matter of preference (whether that be the individual's or the community's). My view counts as objectivist and realist in the sense that it satisfies the following criteria (I will leave open the possibility that there are different shades of realism and mine has a tinge of greyness):

- It does not make wellbeing contingent on preferences, desires, beliefs, or value judgments.
- People can be substantively wrong about their own good, not just mistaken in terms of consistency, but substantively in terms of what contains ultimate (dis)value.

Now to take stock. This discussion of the philosophy of pain was prompted by my intention to defend hedonism, in particular the hedonistic account of disvalue, from EDA. Although the

hedonistic account of disvalue is often deemed to be the view that pain is the only thing of intrinsic disvalue, I have claimed that this is but a sufficient, though unnecessary condition for hedonism, and is not the perspective on *the bad* which I take. I accept a version of the aversion theory of pain, according to which, disvalue lies in the experience of disliking or having an aversion to the sensation of pain, during the time when one is having this experience. This experience is, in and of itself, intrinsically bad and in ascribing disvalue singularly to a certain type of experience, this view satisfies the experience requirement and should rightly be classed as hedonistic. Though I accept that there is still plenty of room for discussion about whether this view is wholly compatible with an objectivist understanding of *the bad*, I believe I have done enough to show that it is objectivist in one important sense: it does not make wellbeing contingent on the beliefs, values, or desires of agents. It therefore meets the requirement of stance-independence which I take to be one of the fundamental tenets of realism.

3. The Evolutionary Debunking Argument Against Hedonism

Hedonists generally take a stance on both *the good* and *the bad*.¹⁰⁸ A full defence of hedonism against EDA will need to show that neither are undermined by EDA. However, I am going to focus primarily on disvalue. This is in large part because evolutionary debunkers – most notably Sharon Street – have paid more attention to how belief in pain’s badness can be explained by evolution. I will therefore leave open the question of whether a similar argument could be used to defend the belief that pleasure is *the good*.¹⁰⁹ Now, strictly speaking, my foray into the philosophy of pain resulted in my denial that pain is intrinsically bad. I claimed that the experience is neutral on its own, but bad when we are hedonically averse to it. This is the view I will be defending against EDA, though to avoid overly cumbersome terminology, I will continue to use phrases like, *belief in pain’s objective badness*, for shorthand for the more nuanced conception of hedonistic disvalue discussed in the previous section. A final reminder of the distinctiveness of this argument: I accept that at some level evolution did shape the belief in pain’s objective badness. Which means that if I can successfully defend this view against EDA, then I will have shown that it is possible for evolution to play a role in shaping a normative belief without undermining it.

¹⁰⁸ Though on some views, such as negative hedonistic utilitarianism, the need to alleviate and avoid disvalue takes precedent over promoting value.

¹⁰⁹ Though it should be noted that since my defence of the belief in *pain’s badness* rests on it being a belief concerning a mental state, it seems likely that such an argument could be applied to *pleasure’s goodness*.

Based on my discussion of the structure of EDA in the second chapter, we can construct the following *local* EDA against the belief in pain's objective badness:

1. You believe that *pain is bad* because members of our species evolved to think this.
2. Since normative knowledge is not fitness enhancing, evolution is not a reliable guide to forming true normative beliefs.
3. If a belief was caused in an unreliable way, then unless there are independent grounds for accepting that belief, we should suspend judgment about whether it is true.
4. There are no independent grounds for believing that *pain is bad*
5. Therefore, we should suspend judgment about whether *pain is bad*

For starters, let's consider the first premise, that the widespread belief in pain's badness was caused by evolution. Pain evolved to motivate humans and other sentient animals to avoid the life-threatening things which pain usually indicates. If people did not dislike the sensations associated with such potentially lethal things, the members of our species would be at more risk of premature death. As Street plausibly suggests:

It is of course no mystery whatsoever, from an evolutionary point of view, why we and the other animals came to take the sensations associated with bodily conditions such as [cuts, burns, bruises, broken bones] to count in favour of what would avoid, lessen, or stop them rather than in favour of what would bring about and intensify them. One need only imagine the reproductive prospects of a creature who relished and sought after the sensations of its bones breaking and its tissues tearing; just think how many descendants such a creature would leave in comparison to those who happened to abhor and avoid such sensations (Street 2006, p.150).

Similarly, Sober and Wilson (1998, p.201) claim that: "There is an evolutionary reason that we have the desire to avoid pain. Pain is typically associated with bodily injury. Present-day organisms seek to avoid pain because this strategy was favoured by natural selection."

These claims seem highly plausible to me. This means that I accept (1). Premise (3) seems fairly unobjectionable on the face of it. Some belief's being caused in an unreliable way should, unless it has independent justification that we know of, be treated with a healthy degree of scepticism. The best strategy for the objectivist about pain's badness will be to object to (2) or (4) or both. My primary focus is going to be on (2), the claim that evolution is an unreliable guide to normative truth; this premise's truth implies that if evolution did shape some normative belief, it most likely was not in the direction of truth. For reasons already outlined, premise (2) is close to the truth, but it should come with the caveat that a very small number of

normative propositions are exempt from it, since our belief in these propositions was caused by evolution in a relatively unique way, such that the anti-realist implications do not follow. The objectivist about pain's badness may now argue:

1. I believe that *being in a state of aversion toward one's pain is bad* because members of our species evolved to think this.
2. Since normative knowledge is not fitness enhancing, evolution is not a reliable guide to forming true normative beliefs, *except for the small number of normative beliefs which bear a particular causal relation to evolution, **Relation R**. Normative beliefs that are R-related to evolution are not, in virtue of this relation, unreliable (though they might be deemed reliable for other reasons).*
3. If a normative belief is *R-related* to evolution, then evolution's causal relationship to that belief does not, in itself, warrant a sceptical attitude toward it (though we may take such an attitude for other reasons).
4. My belief that *being in a state of aversion toward one's pain is bad* is *R-related* to evolution.
5. Therefore, evolution's causal relation to my belief that *being in a state of aversion to one's pain is bad* does not warrant my taking a sceptical attitude towards that belief.

The question that now needs to be addressed is what *Relation R* is and how it differs from the relation between evolution and the normative beliefs it serves to undermine.

3.1. How Evolution Usually Shapes Normative Judgments: Haidt's *Social Intuitionist Model*

Our coming to dislike pain is caused, in large part, by our intimate and direct acquaintance with it in our consciousness. When we are in pain, we do not like it and normally want it to stop immediately. Though I will have more to say about this later, for now, it should just be noted that our belief in pain's badness seems to arise out of a very direct experience of it.¹¹⁰ The way in which we form many other normative beliefs is less direct and is usually facilitated by an intermediary mental state such as an emotion. Many of the philosophers and scientists who endorse EDA draw heavily on the role of *moral emotions*, such as *anger, guilt, shame, disgust* and *sympathy*.¹¹¹ These emotions evolved, it is claimed, to regulate our social behaviour in

¹¹⁰ Along similar lines, Ben Bramble argues that a shared hatred of pain might be a basic trait shared by humans and other organisms, one which evolution has exploited for its own 'ends.' See Bramble (2017). Singer and de Lazari Radek (2014, pp.266-267) indicate a solution to the EDA against hedonism which posits that our reason for believing pain is bad has to do with our experience of pain, but very little detail is provided of what, exactly, this involves.

¹¹¹ A thorough treatment of these emotions is given by Joyce (2006).

fitness enhancing ways. For several proponents of this view, known as normative *sentimentalists* (sometimes *projectivists*), emotions such as these, rather than reason, are responsible for our normative judgments. Joyce, for instance, writes that “[...] if natural selection had a direct hand in shaping the human moral sense, the modification of the brain’s emotional architecture was its principal means” (Joyce 2006, p.124).

The evolutionary psychologist Jonathan Haidt has proposed a model of normative judgment – the *Social Intuitionist Model* – according to which, in most cases,¹¹² normative judgments consist of snap judgments arising from our emotion-based intuitions (Haidt 2001). He gives the example of how people form their views on incest. Haidt suggests that the reason people strongly disapprove of incest is that the thought of it brings about feelings of disgust (ibid, p.814). Haidt adds that people cling to this belief even when they consider scenarios in which such actions bring about no harm. For instance, in their initial rationalisation, Haidt’s subjects pointed to the fact that procreation through incest increased the likelihood of the child being born with a disability. When asked to consider an example of incest in which there is no chance of conception and in which no harm comes to the siblings, people still generally maintain their initial belief. But with their initial defence off the table, they often resort to the *it’s just wrong!* defence (ibid, p.814). The lesson Haidt draws is that reason plays very little role in our normative judgments and is usually confined to a *post hoc* rationalisation of our pre-reflective emotional intuitions (ibid, pp.814-815).

Haidt’s social intuitionist account has been highly influential. In a comprehensive literature review on evolutionary psychological explanations of morality, Haidt’s article came out as the most cited paper (Ellemers et al 2019, p.347). His account has also been influential in the philosophical literature. Richard Joyce draws heavily on Haidt’s account in his book *The Evolution of Morality* in his discussion of moral emotions (Joyce 2006 p.116). In her paper, Street prefaces her argument with the caveat that:

If the evolutionary facts are roughly as I speculate, here is what might be said philosophically. I try to rest my arguments on the least controversial, most well-founded evolutionary speculations possible. But they are speculations nonetheless, and they fall within a difficult and relatively new subfield of evolutionary biology known as evolutionary psychology (Street 2006, p.112).

¹¹² Haidt highlights the normative judgments of philosophers as being a possible exception.

Being highly influential does not necessarily mean that Haidt's account is the least controversial nor the most well founded, but its prominent role in EDAs is reason to consider it a model example of how evolution is supposed to shape normative beliefs.

On the model we are considering, our normative evaluation of some behaviour, whether it be incest, cheating, unfairness or kindness, is facilitated through emotions and these 'moral emotions' have an evolutionary basis. Such emotions play the role of intermediaries when we make normative evaluations, and their presence could be said to *cloud* or *colour* such evaluations, thereby reducing the extent to which we directly perceive the object of our evaluation. It is also worth noting that, of our epistemic access to the objects of our normative evaluation and the emotional mental states through which we evaluate them, it is with our emotional mental states that we have the most direct insight. Suppose that my employer is late in paying me and I cannot afford to pay my bills. This may induce anger and worry which may cloud my consciousness. Though I have direct insight into these mental states, their cause - my employer's lateness in paying me - is not directly present in my consciousness. The *thought of their not paying my check* is within my consciousness, but not the action itself.

3.2. Why the Social Intuitionist Model Does Not Explain Pain's Badness

On the Haidtian model, our conscious access to the objects of our moral appraisal is quite indirect and coloured by the direct intermediary emotion. But there is a different class of normative judgments whose objects are more directly known to us: normative judgments concerning our mental states themselves. These include normative judgments about our emotions, such as the badness of hatred, or the goodness of love. I am focusing on purely hedonic states, in this case the experience of pain, or more precisely, being in a state of aversion toward one's pain. How does this normative judgment, in being about a mental state, differ from normative judgments about non-mental states? One difference is that we are directly acquainted with our pains within our consciousness. Pains can, especially if they are intense, fill our conscious experience. There is no intermediary between ourselves and our pains. Pain is not known via *qualia* for it is a form of *qualia*. And although pain may cause emotion, or its anticipation may induce emotion, when we are actually in pain, that experience need not be accompanied by emotion for us to consider it bad. Unlike non-mental states, we *perceive* our pains – and their badness – clearly and directly.

There might be some objections here. One might be that although pain is not itself an emotion, it is so conceptually bound up with emotions that it is impossible to disentangle them. Meaning

that when we evaluate our pains it is always through the lens of emotion. Now there are two separate questions here. One is, can we experience pain directly and not through an emotional lens? The second is, do we form our normative judgment about pain through an emotional lens? To the first question, my answer is that although pains often accompany and induce emotions, they can be conceptually separated. The pains that can most obviously be separated are brute physical pains such as migraines, being cut or burnt, and tooth ache. It seems at least conceptually possible that one can, even if just momentarily, endure such a pain in the absence of any emotion. Examples showing this are brief unexpected pains whose attendance on our consciousness does not leave enough time for emotion to take hold, such as unwittingly picking up a tray that is scolding hot, or being woken up by an intense pain, such as having a mouse trap affixed to one's toe. Though one might feel emotional following these pains, surely there are some moments, particularly at the very onset of the pain, in which pain alone fills one's consciousness. Another example is all-encompassing pain, such as a cluster headache. No doubt these ailments are the object of negative emotions for those who suffer from them. But when such pains are at their most intense, most other aspects of consciousness will be blocked out by the sheer awfulness the experience.

It follows, then, that we can experience pain directly, in a way that is not mediated by an emotion. I do not think this can necessarily be said about all pains. Mental or attitudinal pains, such as those associated with depression, might be considered hybrid mental states involving emotion and pain. My focus here will be on the brute pains which can be detached from emotional states. With the first question answered, I will now address the second. Even if one can experience pain in the absence of emotion, one might claim that emotion still plays an important role in how we normatively evaluate pain. If I know that I will shortly experience an extreme pain, I might become afraid. If I am in pain and have been for some time, I may find myself in a state of despair. And when I see others in pain, this arouses my sympathy. This raises the question of whether such emotions are an indispensable factor in our coming to believe pain is bad. A reminder of why our answer to this question matters. The evolutionary debunker might claim that although we have a direct apprehension of our pains, our coming to believe *pain is bad* is mediated by emotional states. If this explanation is true, then presumably pain's badness would fall within the bounds of Haidt's social intuitionist model and will thereby, be a prime candidate for evolutionary debunking. On the other hand, if our belief that pain is bad is not mediated through our emotions, then the realist could argue that *pain is bad* is unique far as 'evolved beliefs' go because the relation between the normative belief and the

object of that belief is direct. This is relevant for the following two reasons. The first is that the evolutionary debunker cannot claim that evolution used our moral emotions to ‘mislead’ us into thinking we have a normative reason to avoid pain, thereby depriving them of a key part of their explanation. Second, it would mean that our belief in pain’s badness is based on a more direct perception of the object of our normative belief. At the very least, the realist might claim that in believing pain to be bad we are *seeing things more clearly* because our judgment is not being clouded by emotion, but is based on a direct apprehension of the object of the normative judgment: pain. And it might be argued that this direct insight makes it more epistemically secure than other evolved beliefs which are mediated by emotion.

Is the *judgment* that pain is bad mediated by emotion? I think the right answer lies somewhere between the extremes. To suggest emotion is irrelevant to our belief in pain’s badness is too strong. But to suggest that emotions do all or even most the explanatory work is mistaken. My position is that the *main reason* people believe that pain is bad is not, primarily, because it arouses a negative emotional response, but because they are averse to the way it feels. It is the raw awfulness of the experience, particularly in intense pains, which is the primary causal factor. Suppose one is anticipating some highly painful experience such as an operation performed without anaesthetic. In such a situation, I dread my future pain *because I know the feeling will be awful*. I do not think it is awful *because I fear it*. This gets the relationship completely the wrong way around. The emotion is based on our epistemic insight into the feeling, not *visa versa*. I think the same broadly applies when evaluating other people’s pains. Suppose I see some person wincing in agony, causing sympathy on my part and a desire for their pain to cease. What is going on here? Surely it is not that *sympathy* comes first and leads me to believe the sufferer’s pain is bad. More plausibly, as someone with direct insight into what pain feels like, I recognise that their pain is bad because I know what it is like to have that experience and this knowledge arouses my sympathy.

One thing which raises difficulties for my account is that the language we use to describe pains is emotionally loaded. In objecting to the EDA against pain’s badness, Ben Bramble cites our “shared hatred” of pain as being the cause of our belief in its badness (Bramble 2017). Hatred is a somewhat ambiguous term. It can be used in an emotional sense to describe a mental state akin to anger, whose cause is external to us; e.g., the hate people describe when they assert “I hate Hitler.” Intuitively, this mental state differs from the kind that attends an ascription of hatred to some subjective state, such as the statement that “I hate the taste of kimchi.” Second, one can be said to hate something even when such hate is not attendant upon their

consciousness. For instance, if I hate Hitler, I presumably still hate him even in the moments in which I do not think about him (at least, that's what ordinary language seems to suggest). But when we say we hate pain, our hatred is more in relation to the present experience of pain. Consider the pain of being electrocuted. Bracket off the events which occur in consciousness prior to and post the pain this induces. For now, focus on the actual moments in which the pain lasts. During the moments in which our body is spasming and writhing in agony, we certainly do not seem to have the kind of cognitive-reflective hatred that we do of external things, like when we hate certain people or ideas. This is largely the reason I accept Kahane's contention that the word *aversion* does a better job of capturing the kind of dislike we have of pain. This word has connotations of being raw, primitive and uncontrollable, as our response to pain often is. I also think the word captures the experiential nature of pain's badness. We can "hate" things with which we have had little or perhaps no direct acquaintance, but there is something more *in the moment, direct and experiential* about an aversion. The upshot being, although the word hatred is sometimes used to mean something like an aversion – as I think it is in Bramble's discussion of pain – it also can be used to describe an emotional state. Though as with Parfit, I suspect any difference between Bramble and Kahane on this point is basically semantic.

Several inferences can be drawn from these remarks. The first is that evolution's relation to our belief in pain's badness differs from the way it relates to other evolved normative beliefs, insofar as it is not primarily mediated through an emotional mental state. Furthermore, since these emotions are what, according to debunkers, 'mislead' us into believing in the existence of certain normative principles, the realist can claim that we were not misled into thinking we have a normative reason to avoid pain. That belief arises out of a much more direct apprehension of its object, meaning that we see pain's badness *more clearly*, making it more epistemically secure. Bringing the discussion back to my response to the EDA against pain's badness, we have now identified *relation R*, that unique evolutionary relation whereby evolution's role in shaping our normative judgment does not happen to undermine it, as one where the normative belief was not shaped by an intermediary emotional state. I have argued that belief in the badness of being averse to pain is one such R-related belief. It is likely that the same will be true of brute pleasures, such as the pleasures of eating tasty food and the feeling of warmth on our skin. Whether this argument can be used to defend attitudinal pains and pleasures against EDA is more complicated (see footnote).¹¹³ But attitudinal pains, if that

¹¹³ The distinction between attitudinal and sensory pleasures is a complex one with shades of grey in between. Although there is no 'common-feel' which encapsulates all pains and pleasures, it does seem as though there are general categories which seem more appropriate labels for some pains and pleasures but not others. So called

is what we want to call them, like being in a state of depression, sadness or anxiety are mental states and as such, we apprehend them directly in our consciousness. They are also mental states to which people are averse. This means some of the remarks I have made likely apply to normative judgments about these mental states too.

4. Objections

4.1. Same Debunking, Different Intermediary

According to the most influential account of evolution’s relationship to normative judgments, evolution shapes the content of our normative beliefs via our emotions. I have claimed that an emotional intermediary mental state does not facilitate our belief in pain’s badness; this belief is based upon our aversion to the sensation. However, it could be argued that this argument amounts to little. All I have done is change the intermediary state by which evolution shapes our normative judgment. On the social intuitionist account, this role is fulfilled by an emotional state, but it need not be. It could be argued that in the case of pain’s badness, evolution facilitates this belief via the intermediary state of a *disliking* or *aversion*. If so, the main point stands. We believe pain is bad because evolution has caused us to dislike this sensation – and we dislike this sensation, ultimately, because avoiding pain has survival value, not because pain is actually bad for us.

My response to this point demonstrates the need to take a view on what, precisely, we mean when we claim that pain is bad. If I claim that the sensation of pain alone is bad, then the argument we are considering has considerable force. But I have denied that pain is intrinsically bad. I claimed that the experience of being in a state of aversion toward one’s pain is bad. It is not clear to me how the above objection deals with this view. It surely cannot be that my belief that *being in a state of aversion toward my pain* is bad because I am averse to being in a state of aversion toward my pain. For illustrative purposes, consider the following:

Object of normative belief	Intermediary Mental State	Normative belief
Incest	Disgust	Incest is wrong
Pain	Aversion	Pain is bad

sensory (or bodily) pains usually denote experiences like being burnt, headaches and so on. Attitudinal (or mental) pains typically denote experiences such as depression and sadness. The distinction is, of course, misleading insofar as all pains are ultimately mental experiences. Perhaps the issue here is the word “pain.” Being burned is clearly painful. Is depression painful? Depression diminishes wellbeing and is *bad for us*, but it does not seem to be painful in the sense that being burnt does. That does not make it any less bad, of course. As I see it, what is bad according to the aversion theory is being averse to a particular conscious experience. Many are averse to sensory pains, but people are also averse to attitudinal states like depression, anxiety and sadness. So, intuitively, though I have not argued this here, such experiences should be covered by the argument of this chapter.

Being in a state of aversion toward my pain	Aversion	Being in a state of aversion toward my pain is bad
---	----------	--

I accept that, in the first two cases, the intermediary mental state plays a role in leading us to the normative belief. But it makes very little sense to say this about the third case. I claim that the mental state of being averse to some present pain is of negative value. But it makes no sense to add an additional layer of aversion as an intermediary between this experience and my belief that it is bad, given that the mental state being ascribed disvalue is a state involving aversion. More plausibly, the move from this mental state to our judgment is more direct:

Object of normative belief	Normative Belief
Being in a state of aversion toward my pain	Being in a state of aversion toward my pain is bad

4.2. The Contingency of Pain's Badness

For reasons outlined in the first chapter, realists hold that normative truths are necessary, not contingent on the stances of agents. Sharon Street is sceptical about this being true of the belief in pain's badness, prompting us to consider a hypothetical counterexample.¹¹⁴ Suppose that on some other planet, an alien species evolved in conditions radically different to ours, such that there was no evolutionary advantage to be gained from caring about pain. Such beings would presumably have a radically different response to pain than ours. If their survival did not rest on the avoidance of pain, perhaps they would exhibit no negative response to it. If so, it would be difficult to see why their pain would be bad for them. This suggests that whether pain is bad is contingent on our evolved response to pain, rather than there being something necessarily bad about it. The fact that humans consider pain to be bad just tells us something about their evolutionary history. If we evolved to be neutral about the sensation of pain, we probably would

¹¹⁴ In Street's (2006) paper, she argues against a realist construal of the *badness of pain* on the grounds that it is contingent on our evolved evaluative responses to pain (pp.144-152). In the end, her argument seems to presuppose the realist must be committed to some kind of sensation theory of pain, which Street rejects. But for reasons outlined earlier I think this claim is mistaken and that realism is compatible with dislike/aversion theories. Street's thought experiment about creatures who are indifferent to pain can be found in Street (2009a, pp.287-288). In this paper, Street is responding to Parfit's *Future Tuesday Indifference* scenario in which a person cares about their pains except those which occur on Thursdays (Parfit 2011a, pp.56-57). Parfit claims that such a preference is objectively irrational. Street's thought experiment is tailored to deal with this objection, but I have generalised it in such a way that the aliens are simply indifferent to their pain irrespective of when it occurs.

not hold that view. If pain's badness is contingent in this way, perhaps we should be anti-realists about pain's badness.

I accept that such aliens are unlikely to form the judgment that their pain is bad for them. Nor should we humans believe such aliens are mistaken about that claim, given that their ability to feel pain puts them in a good epistemic position from which to make that assessment. However, these aliens are not in a sound epistemic situation to judge whether *being in a state of aversion toward their pain* is bad. Why? Because they do not respond to their pains in this way. One cannot, for example, claim that these pain-indifferent aliens are indifferent about *being in a state of aversion toward their pain* because that mental state is by its very nature not indifferent. Moreover, let's suppose that, by some extravagant technology we were able to provide the aliens with this experience by some artificial means. That is, through manipulation of their nervous system, we are going to give the aliens the experience of – depending on your preferred terminology – *being in a state of disliking, hating, or being averse to their pain*. Would that mental state be bad for the aliens? I cannot make a determination with *absolute* certainty. But intuitively, the thought that this would harm the aliens has more force, as I see it, than its negation.

A good way of testing our intuitions on this would be to consider whether it would be ethical to subject the aliens to what humans perceive when they are in a state of being averse to their pain, e.g., the feeling most humans have when being electrocuted. Intuitively, I think it would be wrong to perform such an experiment precisely because I think it would involve harming the alien (I'll leave it to the reader to scrutinise their own intuitions on this case). If you share this intuition, you have strong grounds to reject an evolutionary debunking of the claim that *being in a state of aversion toward my pain* is bad, because these hypothetical aliens did not evolve to care about this experience. In which case, if that mental state is bad for them, it is solely because of its phenomenological quality. Moreover, in making this claim, I am not claiming that *the bad* is contingent. It is contingent (on things such as whether we are pain asymbolics or pain-indifferent aliens) as to whether *pain is bad*. But I have not claimed that pain is intrinsically bad. I claim that being in a state of aversion toward pain is intrinsically bad. And this claim is not contingent. Being in a state of aversion toward one's pain is bad in all possible worlds, including those in which organisms evolved not to care about pain or aversion.

Consider again the proposed evolutionary explanation of pain's badness. We hold pain to be bad, it is claimed, because this makes us more likely to avoid the things which this aversion indicates. This seems reasonable, but it does raise the question, what is so special about the aversion we have toward pain?¹¹⁵ If being averse to pain plays this purely instrumental role, and if there is nothing *intrinsically significant* about this experience, why was it specifically this kind of aversion which evolved to motivate us to avoid harm? Why could a similar motivation not be derived from our perceiving *the colour mauve, the sound of pitter patter, or the smell of strawberries* and to have such experiences in the presence of being burnt, bitten or frozen instead of pain, to get us to avoid such life-threatening events? The fact that aversions to pain phenomena are so effective in getting organisms to avoid harm indicates to me that there is something intrinsically significant about this experience. And in this respect, I tend to agree with Bramble (2017) that the "shared hatred" we have of pain is basic and evolution uses this to its advantage. And that is why it makes perfect evolutionary sense that this is the experience, rather than perceiving mauve, the sound of pitter patter, or the smell of strawberries, that accompanies those life-threatening things which it is in our evolutionary interest to avoid. Otherwise put, being in a state of aversion toward pain is so effective in getting organisms to avoid certain harmful situations because that experience really is bad for us and one we have reason to avoid.

A final worry might be that my account is somewhat muddled. I have given a fairly indirect evolutionary account for how we can recognise the truth of utilitarian claims associated with *the right* but I have now suggested evolution has much more direct involvement in our ability to recognise the truth of hedonism. Why the asymmetry between our knowledge of *the right* and *the good*? My response is that I see no reason why we need to give a similar explanation in both cases. *The good* pertains to considerations of wellbeing; to the question of what makes our lives go well or badly. As I see it, our experiences are central to answering this question. Knowing what pain feels like is indispensable to knowing whether it enhances or diminishes our wellbeing. Claims about *the right* seem to me to be different to some degree - if we are utilitarians, we think that the right action is to promote *the good* and belief in hedonism is also based on the arguments in favour of it - but different in the sense that questions of distribution do not as obviously require introspection into the nature of our experiences. Indeed, it seems to me that the utilitarian intuitions behind maximisation and impartiality – that there is no "magic [...] in the pronoun "my," that should justify us in overturning the decisions of impartial

¹¹⁵ For more on this point, see Bramble (2017, pp.98-99).

truth [...]” (Godwin 1793/2001, p.82) as Godwin put it - are of a more rationalistic, almost mathematical nature, than our belief in pain’s badness.¹¹⁶

Conclusion

It was not inevitable that we came to believe in pain’s badness. If we were subjected to radically different evolutionary pressures, we might have evolved to be neutral about our pains. We evolved not only to feel pain, but for the most part, to be averse to it. It is for this reason that I accepted Street’s claim that our belief in pain’s badness can in fact be explained by our species’ evolutionary past. But that does not mean this belief can be undermined by EDA. The problematic relation involves evolution shaping some normative belief via some intermediary emotional state. But this does not apply to our belief that being in a state of aversion toward pain is bad, for this belief was not shaped by our emotions, but by our aversion to the way it feels. I have now completed my defence of hedonistic utilitarianism against Sharon Street’s content partial EDA and in the process, shown that 1) some normative beliefs were most likely not shaped by evolution and therefore are immune to EDA and 2) at least one normative belief – belief in pain’s objective badness – was shaped by evolution, but not in a way which undermines it (though I have conceded that this belief may be quite unique in this respect). In the next and final chapter, I will respond to the content-neutral variant of the EDA and provide some methodological and epistemological remarks on how we can go about acquiring normative knowledge notwithstanding the distortive role evolution has had on our normative thinking.

¹¹⁶ An analogy between Sidgwick’s axioms and mathematical knowledge is made by Singer and de Lazari-Radek (2014, p.120).

Chapter 5: Where Does Normative Knowledge Come From?

Introduction

The two preceding chapters consisted of responses to Sharon Street's *Content Partial* EDA. In the first part of this chapter, I address the *Content Neutral* EDA developed by philosophers such as Richard Joyce and Phillip Kitcher. According to this argument, the act of thinking normatively and our use of normative concepts, are products of evolution designed to solve collective action problems and maintain social cohesion. Our belief that normative reasons exist is ultimately a useful fiction - a "myth," according to the title of one of Joyce's books on this subject - that helped our ancestors to survive and reproduce.¹¹⁷ By way of response, I develop a rival normative genealogy showing how normative knowledge is possible. I accept that some aspects of our normative thinking originate in evolution, particularly those gut-based, instinctive normative beliefs which tend to characterise common-sense normative thinking. However, there is a class of normative judgments of a more reflective nature whose origins lie not in our instinctive emotional responses, but in our ability to engage in critical thinking, introspection, and reflection. In this respect, my genealogy is dual-aspect, and the problem with the content-neutral EDA is that it can only explain one of these aspects.

In the second part, I address issues pertaining to normative epistemology and methodology. Thus far, I have established that although evolution has not shaped all our normative judgments, it has shaped many. This raises questions about how we can differentiate between beliefs which evolution shaped and those it did not. Identifying an appropriate method for this task is especially important if we accept that evolution's influence is a distortive one which tends to shape our normative judgments in directions unrelated to truth. Perhaps the most widely used normative methodology is reflective equilibrium.¹¹⁸ But this approach is arguably inadequate since it places too much emphasis on testing prospective normative principles against our background normative intuitions. If our background intuitions are tainted by evolution's influence, then they are possibly not a suitable guide toward truth. I will outline a foundherentist methodology which includes insights from both foundationalism and coherentism to deal with these issues.

¹¹⁷ Joyce has developed a view known as moral fictionalism. On this view, no moral claims are true. But Joyce maintains that we should continue to use moral concepts for the practical benefits they bring, but to treat them for what they are: useful fictions. For Joyce's account of fictionalism, see Joyce (2001, chs.7&8).

¹¹⁸ Reflective equilibrium was first outlined by John Rawls in *A Theory of Justice* (1973).

1. Normative Genealogies

A normative genealogy seeks to identify the origins of our normative ideas, concepts and beliefs. Though a worthwhile project in themselves, genealogies are often aimed at establishing how the origins of our normative thinking have further implications concerning the nature and justification of our normative beliefs. So-called *Genealogical Debunking Arguments* claim that if some normative belief has dubious origins, we should take a sceptical stance towards it. EDA is one species of genealogical debunking argument. But others have been outlined identifying non-evolutionary origins as having a questionable influence on our normative thinking.¹¹⁹ Street's evolutionary genealogy is content-partial insofar as it asserts that the content of our normative beliefs was shaped by the forces of natural selection. Joyce's genealogy is content-neutral insofar as he does not seek to explain the content of our normative beliefs, but the reason we think normatively in the first place.

Street and Joyce also disagree regarding what they take to be the implications of their respective genealogies. For Street, the evolutionary origin of our normative beliefs implies we are unlikely to ever encounter the stance-independent normative truths posited by the realist. However, Street maintains that such evolutionary origins are compatible with her anti-realist constructivist understanding of normative facts, according to which normative truths are contingent on what agents value (Street 2006, pp.152-154). Joyce's conclusion is more radical. He claims that the evolutionary origins of normative thinking counts against the existence of normative truths of any kind, including anti-realist conceptions. Joyce also claims that his genealogy implies a particular "epistemological conclusion: that all moral judgements are unjustified" (Joyce 2013, p.351). It is in this respect that Joyce is an error-theorist about normativity; or at least the moral aspect of normativity. In his view, our normative beliefs are systematically false because they fail to refer to anything.

Indispensable to Joyce's account is the notion of *categorical imperative* (Joyce 2001, pp.138-140). A categorical imperative to act is one that I have regardless of my personal preferences. If I have a categorical reason to *x* then I ought to *x* even if I do not want to. Response-dependent metaethicists might deny that such reasons exist if they believe that our reasons are contingent on our desires.¹²⁰ Joyce's point, however, is that our *folk normative epistemology* seems to fit with the idea that there are categorical reasons. Most think that we have a reason not to assault

¹¹⁹ An example being Nietzsche's *On the Genealogy of Morals* (1887/2012).

¹²⁰ See Williams (1981, ch.8).

random people regardless of whether we want to. What is the point of believing in categorical reasons? The issue again boils down to the selective pressures that govern existence in social communities. If everyone just went around stealing, free-riding or committing various other anti-social acts whenever it suited their personal preferences, social harmony would soon break down. A sense that certain things ought to be done irrespective of my preferences, facilitated by the motivating force of our normative emotions, ensured that members of the group acted in pro-social ways even when they did not desire to do so.

1.1. An Outline of a Dualist Normative Genealogy

Joyce's argument asserts that the realist's belief in stance-independent normative facts is basically a delusory one which we find compelling because our species has benefitted from believing that certain things ought and ought not to be done irrespective of our personal preferences. I will not be providing a wholesale rejection of Joyce's genealogy because I think it is partially correct. The problem is that it is too limited. It correctly characterises one aspect of our normative thinking, in particular, our pre-reflective instinctive normative judgments. But it is less successful in characterising our more reflective normative judgments. The normative genealogy I favour is two-pronged. It seeks to explain both the reflective and instinctive aspects of our normative thinking by attributing these distinctive aspects to different sources. By offering a dualistic normative genealogy I do not mean to suggest that the two sources I cite are exhaustive of the origins of normativity. Rather, I am arguing that there are at least two aspects to normative thought, and I leave open the question of whether there are more.

1.2. The Role of Evolution

In chapter two, I summarised some of the literature in evolutionary psychology and primatology arguing that human normative thinking has its origins in evolution. Some of the behaviours of our close evolutionary relatives, in particular Chimpanzees and Bonobos, seem to resemble human normative practices. This all seems to fit with the idea that normative thinking evolved to deal with problems arising out of social living. Indeed, I would even go one further than Joyce and suggest that such data most likely explain the content of many widely held beliefs, *blood is thicker than water*, *do unto others as you would have them do unto you* etc.

This is without mentioning the way in which non-human primate normativity is most like human normativity: the importance of emotion. Although I have never experienced being a chimpanzee, it is uncontroversial that they and other non-human primates feel emotions in response to certain behaviours and that these emotions play a motivating force in the policing of norms (e.g., the anger prompting the brown capuchin to throw their unequal share of food out of the cage or the sympathy driving a chimpanzee to console the victim of an attack). There is clearly a difference between being a moral being and an emotional being. If some organism spends their entire existence in a constant state of sorrow, we might call them an emotional being, but not necessarily a moral one. Nonetheless, according to many moral philosophers and evolutionary psychologists, so-called moral emotions are an indispensable part of normativity. Not only that, but our emotions are the *prime mover*. Jonathan Haidt, whose highly influential account of normative judgment I discussed in the previous chapter, entitled his paper “The Emotional Dog and Its Rational Tail,” a metaphorical description of normative judgment long held by Humean philosophers who believe that “Reason is, and ought only to be, the slave of the passions” (Hume 1985/1739, bk.3, pt.3, S.3).

One should not exaggerate how much these primatological accounts tell us about human normative thinking. Human normative practices differ in important respects. We might even ask whether it is plausible to say that Chimpanzees and Bonobos *think* normatively. Questions like this are impossible to answer with certainty given the epistemological barriers involved in interpreting the mental states of other species. It seems likely that non-human primates do not hold propositionalised judgments (e.g., a belief of the content *it is wrong to murder*). This arguably would require the use of language. Similarly, it is unlikely that non-human primates possess an understanding of normative concepts: the *idea* that something *ought to be done*, *is fair* etc. But this does not mean non-human primates cannot think normatively. Indeed, I would suggest that the consistency with which chimpanzees punish certain forms of behaviour and their apparent recognition of certain social norms and expectations demonstrates normative thinking of a kind. At the very least, it shows something akin to taking a normative disposition toward certain types of acts and behaviours. That chimpanzees and bonobos hold normative beliefs seems to me much more doubtful.

I now want to consider to what extent Joyce’s genealogy can explain human normativity. It provides quite a convincing account of the origin of our ‘moral emotions.’ It is not just by some coincidence that we tend to get angry when we observe someone free riding and sympathise with the victims of indiscriminate violence. For organisms reliant on survival within a social

setting, such as primates, these emotional responses served an important function. However, having emotional responses to certain behaviours and acts falls short of characterising human normativity even if it forms an essential part of it. It leaves out our ability to construct propositionalised normative statements and the use of normative concepts. This is not necessarily a problem. What the primatological studies suggest is that there is some ancestor common to humans and chimpanzees that possessed the building blocks of normativity. In the case of our evolutionary cousins these building blocks have not progressed further towards human normativity because these primates did not evolve the requisite traits, such as the use of language. It does not seem too much of a stretch to suggest that one of the selective pressures culminating in our use of language is the role it plays in the transmission of norms and the education of group members in the standards of behaviour which contribute to social cohesion and stability. The building blocks of normative thinking predated human beings, but our advanced linguistic and reasoning capabilities allowed us to take it to new levels. It also allowed us to engage in a more sophisticated form of moral reasoning whereby we make inferences based on our basic normative judgements (e.g., *x* is wrong; *f* involves *x*-ing; therefore, *f* is wrong).

1.3. Where Joyce's Genealogy Falls Short

I have noted some of the similarities between human and non-human normative behaviour and conceded that there is evidence that evolution played a role in human normative thinking. I discussed one dissimilarity between the two, namely the human ability to form propositionalised normative beliefs, but noted that this does not undermine Joyce's genealogy since it just shows one way that humans were able to take normative practice further than their evolutionary cousins and the common ancestors they share with them. There are, however, more striking differences between human and non-human beings that point to an aspect of our normative thinking that is more difficult to reconcile with Joyce's genealogy.

Non-human primate normativity consists primarily of instinctive reactions to certain types of behaviour. What chimpanzees and bonobos do not do is ask questions about *how we ought to live*. A chimpanzee might implicitly recognise a social expectation to reciprocate altruism and act accordingly by punishing non-reciprocation. But so far as we know, chimpanzees do not reflect and deliberate about whether their society's endorsement of these norms is *justified*. It might be thought that this is a trivial point. Chimpanzees cannot reflect on norms in this way

because they do not possess the necessary intellectual abilities. But this is precisely my point and it is far from trivial. I accept that like non-human primates, human beings evolved to observe, accept and reject certain types of behaviour because it contributed to our ancestors' survival. This aspect of our normative thinking, one that involves following and enforcing norms based on gut emotional instincts, can be explained by evolution. But blindly accepting a set of norms is only one side of the coin. Humans are uniquely the animals that question the norms of their societies and search for new ones.¹²¹

I now want to say more about this *questioning* aspect of our normative thinking. Though many people unquestioningly accept the dominant norms within their societies, it is an empirical fact that in most (probably all) cultures, there are people who have taken a questioning attitude toward the dominant societal norms. Rather than acquiescing to societal prescriptions, they ask themselves the Socratic question of *how we ought to live*. Indeed, Socrates is a prime example of a person who embodies this *questioning spirit*. Not only do humans often take a questioning stance toward their society's normative values, but they question their own pre-reflective normative beliefs. The Socratic dialogue discussed in chapter three is instructive here in which Socrates and Crito are questioning the widely held judgment that those who commit wrongdoing deserve punishment. The view we take on this matter may depend on the context in which we find ourselves. If I am the victim of wrongdoing – let's say that a gang of thieves has broken into my home and stolen all my belongings – my initial response is likely to be vengeful. Having reflected on the situation in a less impassioned way, my response may be more measured, though I will still probably be resentful. But I can also think about this issue in a more impersonal way. Reflecting in an abstract philosophical way, I can consider whether we are justified in believing that wrongdoers ought to be punished. I can consider whether our grounds for believing in retributive punishment stand up to critical reflection.¹²²

Irrespective of whether I respond to wrongdoing in a retributive manner in situations where I am the victim, reflecting in a calm and sober manner I might come to the belief that retributive punishment is unjustified. As it happens, this is the view I hold, though I recognise that in certain situations I might act contrarily (say, if I am personally the victim of some serious

¹²¹ A similar line of argument is developed by Korsgaard in response to de Waal's citation of the behaviour of primates as evidence for an evolved morality. See Korsgaard (2006).

¹²² Singer and de Lazari-Radek, in their discussion of Sidgwick's ethics, also note the importance of distinguishing between instinctive and more reflective moral beliefs. They see Sidgwick's axioms as having their origins in our capacity to reason, a claim which forms the basis of their response to EDA against the principle of universal benevolence. See Singer and de Lazari Radek (2014, ch.7).

wrongdoing). Likewise, I hold the utilitarian view that the right action is to maximise welfare impartially. Though I suspect if I were to win a sizeable amount of money on the lottery my gut reaction would be to keep most of the money and distribute a sizeable chunk to my close relatives; a gut reaction which fits with an evolutionary picture of human nature, but conflicts with the normative theory I accept. This illustrates that our normative thinking is sometimes instinctive and unreflective (my retributive response to the person who wrongs me) and sometimes reflective and thought-through (engaging in a philosophical discussion or reflection on the question of whether wrongdoers deserve to be punished). These are the two aspects of human normative thinking. Though in some cases they proceed harmoniously, in other cases they can pull us in radically different directions.

Joyce's genealogy is too limited, because although it explains our pre-reflective normative beliefs, it does not provide an adequate explanation of our more reflective normative beliefs which often conflict with our instincts. From where do these normative beliefs originate? I sketched an answer in the previous chapter, where I discussed rationalist views which claim our ability to discover normative truths is a by-product of our general reasoning capabilities. However, as it stands this explanation is too vague and requires more detail. Consider again the example of Socrates. Socrates questioned the basic beliefs many within his society took for granted. In considering prospective answers to some puzzling question, he would attempt to identify counterexamples and considered the logical consequences of the prospective hypothesis to work out whether they conflict with our wider beliefs. This intellectual activity is the essence of *philosophy*.

Russ Shafer-Landau has argued that ethics is a species of reasoning whose genus is philosophy (Shafer-Landau 2006). The hallmarks of philosophical inquiry are all present in ethical inquiry. Here are some of the similarities Shafer-Landau mentions. First, neither philosophy nor ethics seem to be branches of the natural sciences insofar as neither are primarily engaged with enumerative induction nor with providing explanatory descriptive hypotheses. Providing a descriptive explanation for why people believe *x is good* is a job for the scientist; the philosopher is concerned with whether *we ought to believe x is good*. Likewise, why we believe in an external world is a descriptive question; whether we ought to believe in an external world is philosophical. Moreover, the objects of philosophical and normative inquiry both seem abstract. One cannot observe *oughts* and *duties*, these are non-natural, abstract concepts that supervene on natural concepts. But the same is true of the objects of our philosophical inquiry. Shafer-Landau notes that "Philosophy is not primarily an empirical discipline, but an a priori

one. Its truths are ordinarily discoverable, when they are, not exclusively by appeal to what our senses can tell us. We don't bump into such things as universals, free will, or modalities; we can't see them, or hear or touch them" (Shafer-Landau 2006, p.216-217).

Our ability to engage in philosophical reasoning can be applied to a range of abstract questions such as 'what is knowledge?' 'does free will exist?' and 'how ought I to live?' Consider the similarities in how we attempt to answer these questions. We might start with an intuitive hypothesis, e.g., the belief that knowledge is justified true belief. We then seek for counterexamples to the hypothesis, a search which may result in our discovery of something surprising, such as Gettier's apparent discovery that one can have a justified true belief which does not qualify as knowledge (Gettier 1963). This type of reasoning is often employed in our consideration of an intuitive normative hypothesis. In considering the view that *the good* is the satisfaction of our desires, the normative ethicist might point to hypothetical situations borne out of thought experiments in which this belief has counterintuitive implications. E.g., If I desire to diminish my wellbeing, then the satisfaction of this desire both promotes and demotes my wellbeing.¹²³ The philosopher will check to see whether the hypothesis is internally consistent with their other beliefs. And as Shafer-Landau notes, in both philosophy and ethics "we would have to investigate to see whether [our hypothesis], among competitors, best exemplifies a host of theoretical virtues—economy, stability across cases, avoidance of ad hoc assumptions, preservation of existing beliefs, explanatory breadth, etc." (ibid). As I see it, it is no accident that ethics is generally considered a branch of philosophy since the tools of philosophy are well suited for addressing questions within that domain.

I just want to pause here to make a clarification. In claiming that one dimension of our normative thinking emanates from our ability to engage in philosophical reasoning, I am not making the elitist claim that academic philosophers have some special insight into an aspect of normativity from which others are excluded. Academic philosophy is a refined and rigorous application of the critical reasoning that humans apply to a range of issues. The basic distinction I am trying to draw in my genealogy is between those normative beliefs which emanate from 'the heart' and those from 'the head'. Some of our normative beliefs seem primarily to be based upon our pre-reflective instinctive normative beliefs, others come to us by a process of reflection, deliberation, or argumentation. Some, no doubt, are a combination of both.

¹²³ For a discussion of an argument of this kind against the desire-fulfilment theory of wellbeing, see Bronsteen (2017).

The proximate cause of our reflective normative beliefs is our ability to engage in critical reasoning of a kind exemplified in the practice of philosophy. I say “proximate” cause because this ability has prior causes also. I do not know how it was that human beings acquired the ability to engage in philosophy. Evolution obviously played a role in providing us with linguistic and reasoning capabilities. I do not think that an ability to discover philosophical facts per se was directly selected for. More likely this ability came about as an indirect consequence of other reasoning abilities that were. Historical cultural changes most likely played a role as well. For our purposes, answering this question is not important. Enough has been done to establish the outline of my dualist normative genealogy. On the one hand, there are pre-reflective normative beliefs and attitudes that evolved to promote social stability and cohesion. Then there are the more reflective normative beliefs that originate in our ability to engage in critical reasoning. The latter often being the product of a *search* for justified norms rather than an instinctive acceptance of norms. Just going back to the issue of disagreement discussed in chapter three, the truth of this dualistic genealogy might hold some explanatory power. Our pre-reflective normative thinking might have resulted in us believing that n , whereas the more critical and reflective aspect of our normative thinking might result in our belief that $\sim n$. This would be no surprise given that both beliefs arose from different origins.

This dualistic normative genealogy is not just the product of arm-chair reasoning. There is empirical evidence that our normative judgments come from different aspects of our cognition. Joshua Greene is a neuroscientist with a background in philosophy who specialises in the psychology of moral judgments. According to his *dual process* model of moral thinking, moral thinking consists of both snap emotional judgments and slower, more reason-based moral judgments (Green 2014, ch.5). Greene performed brain scans on subjects as they considered the notorious *Trolley Problem*¹²⁴ thought experiment (ibid, ch.4). In the ordinary trolley problem, subjects consider whether it is permissible to divert a trolley heading for a group of people into the path of a single person. In the so-called *footbridge* variation, they consider whether it is permissible to push a large person or backpacker from a footbridge into the path of the trolley to prevent it from running over a group of people. In general, most people tend to hold utilitarian intuitions (save the many over the few) when considering the original trolley problem whereas in the footbridge case, most people hold deontological intuitions, believing it is wrong to use the innocent backpacker as a mere means to save those on the track.

¹²⁴ The Trolley Problem was first discussed in a paper by Phillipa Foot (1967).

What is more striking is that Greene's brain scans showed that certain kinds of judgment correlate with activity in certain regions of the brain (ibid, pp.121-124). Those subjects who gave deontological responses typically showed heightened activity in the ventromedial pre-frontal cortex and the amygdala; parts of the brain associated with emotion. Whereas utilitarian responses typically correlated with activity in the dorsolateral pre-frontal cortex, the part of the brain associated with critical reasoning. It is also noteworthy that utilitarian judgments typically took longer to make, again suggesting they are based more on rational calculation as opposed to instinct (ibid, p.127). According to Greene, this dualism reflects the fact that it would make evolutionary sense in many cases for our normative judgments to be made quickly. It would not be very efficient for organisms to reflect on every moral decision they make. Whereas in other cases, a more slow-paced, rational approach to decision-making would be more appropriate. Greene applies this logic to the asymmetry in responses to the ordinary and footbridge trolley-problem cases. In response to the ordinary trolley problem, most people are utilitarian; in the footbridge case, most think deontologically. Depending on one's first order normative view, there may be salient moral differences between the two cases. The ordinary trolley problem seemingly does not violate anybody's rights, whereas the footbridge case involves the initiation of violence against the innocent backpacker and violates the Kantian principle that it is wrong to use someone as a *mere means*. Greene suggests that most people's asymmetrical responses have an evolutionary explanation (ibid, ch.9). The footbridge case involves the initiation of physical violence against the backpacker. Humans have an evolved aversion to physical violence which explains why a deontological judgment in this situation shows activity in the parts of the brain associated with emotion. Why do many of the same people hold utilitarian judgments in response to the ordinary trolley problem? Because unlike physical violence, we did not evolve any responses to levers and trolleys, which are recent features of human history. Greene's experiments bode in favour of the rationalist's claim that some of our normative judgments are reason-based; specifically, those which are of a utilitarian nature.

1.4. Implications for Realism vs Anti-Realism

According to Joyce, belief in objectively binding normative principles was useful in evolutionary terms but has no basis in fact. My response is that Joyce's genealogy is too limited. It only covers our instinctive, emotion-based judgments but not those grounded in critical reflection. One response might be that this duality is irrelevant. Any anti-realist implications concerning the ancient aspect of our normative thinking, must necessarily spill over into the

more reflective dimension. However, I see no reason why this is so. Our ‘folk scientific’ beliefs might have been very useful in helping our ancestors to survive in their environment. But if we find that these beliefs conflict with new findings in quantum physics, it would have no bearing on the status of quantum physics. Likewise, even if Joyce is right that our folk normative beliefs can be debunked by their evolutionary origins, this would not show that our more reflective and reason-based normative beliefs could be debunked in the same way. Joyce would need to provide a separate debunking argument aimed at those normative beliefs whose origins lie in our ability to reason.

2. Normative Epistemology

Epistemology is the branch of philosophy concerned with knowledge. Assuming there is an external reality beyond our existence, how can we come to know about it? How can we tell whether our beliefs are true or false? What is the structure of our knowledge? How can we be sure our beliefs are not distorted by factors such as evolution? These questions can be applied to the normative domain too. As a realist, I think that one aim of normative reasoning is to discriminate between normative truth and normative falsehood. But a satisfactory answer to how we do this will need to address the question of how normative knowledge is acquired.

2.1. The Structure of Normative Knowledge

How do we go about justifying our normative beliefs. One widely held view of epistemic justification is *Coherentism*. According to coherentism, a prospective belief is justified to the extent that it coheres with our wider set of beliefs (Haack 2002, p.418). If a prospective normative belief coheres with our wider normative beliefs, then it is *prima facie* justified. If it does not cohere, then that belief is not justified. An alternative account of justification is *Foundationalism*. According to the foundationalist, the structure of knowledge is vertical. It will not do to proceed by establishing coherence because we first need to be sure that our background beliefs are justified (ibid). Hence the foundationalist works from the ground up, having established secure foundations. We first need to identify some basic propositions that are self-standing, in that they do not derive their justification from other beliefs. It is this search for a self-standing belief that led Descartes to the Cogito: even if all my background beliefs are mistaken, I can be sure that this current act of thinking implies my existence.¹²⁵ Having

¹²⁵ Descartes makes this argument in the second of his “meditations”.

identified the basic building blocks of knowledge, the hope is then that we can build on these secure foundations and account for our wider set of beliefs.

Following the publication of *A Theory of Justice* (1973), John Rawls's method of *Reflective Equilibrium* has become perhaps the most widely endorsed means of justifying normative principles. Reflective equilibrium follows a coherentist structure. To establish whether some normative belief *n* is justified, we seek to ascertain the extent to which *n* coheres with our background beliefs (Rawls 1973, pp.48-51). If we find that *n* conflicts with our most justified background beliefs, this counts against believing that *n*. However, prospective normative principles are not always subservient to our background normative beliefs. Reflective equilibrium also has a revisionist component.¹²⁶ Suppose *n* has considerable intuitive force, but although it coheres with most of our background normative beliefs, it conflicts with a small number of them. For example, suppose someone is considering the principle that *abortion is permissible*. Let's suppose that although most of their background beliefs do not conflict with this principle, one of them does: their pre-existing belief in *the sanctity of human life*, according to which it is *always* impermissible to end a human life. Although this principle conflicts with one of this person's background beliefs, reflective equilibrium does not necessarily imply that the prospective principle being tested needs to be rejected. However, should they adopt this principle whilst maintaining belief in the sanctity of human life, there are problems concerning coherence. This means that a process of reflection will need to take place whereby the person decides which of the two principles is most compelling, or whether either principle can be amended in such a way as to establish coherence. By doing so, a reflective equilibrium will be achieved.

2.2. Problems with Reflective Equilibrium

Reflective equilibrium faces several objections. Several of these are general objections against coherentist theories of justification.

Objection 1: The Infinite Regress

When asked why we believe some proposition, we often appeal to other beliefs. For instance, if asked why I believe the sun will rise tomorrow, I might reply that it has risen on every previous day of which I am aware, and this is good evidence that it will arise tomorrow. When

¹²⁶ A distinction is often made between so-called wide and narrow reflective equilibrium. Wide reflective equilibrium is more revisionist insofar as it allows for more revision of our background beliefs if we find some normative hypothesis compelling. For an account of wide reflective equilibrium, see Daniels (1996, pp.1-47).

asked why I believe that this form of inductive reasoning is justified, I will soon start to struggle in my appeal to pre-existing beliefs. Whether some n is justified, according to the coherentist, will depend on whether it is supported by our wider set of beliefs. In claiming this, they deny the foundationalist's claim that some normative beliefs are self-supporting on the grounds of their self-evidence. The issue is that if there are no self-standing normative beliefs, our chain of justification will never come to an end. Or insofar as it does end, it does so because we have ran out of background beliefs to appeal to (Haack 1993, pp.21-22). The foundationalist will claim that this means coherentism ultimately leads to scepticism. To alleviate these sceptical worries, we need to appeal to the self-evidence of some self-standing normative belief.

Objection 2: The Coherent Fairy Story

The *coherent fairy story objection* calls into question the idea that coherence can provide a justificatory role. It is not logically impossible to imagine a person who has a perfectly coherent set of false beliefs (Haack 1993, pp.26-27). A master logician could conceivably construct a wild conspiracy theory that is wholly internally consistent. If this ultra-logical conspiracy theorist were to find that an outlandish belief, such as the belief that the King of the United Kingdom is a reptilian, coheres with their false world view, then according to the coherentist, they would have good epistemic reason to accept that belief. But presumably, given the ridiculousness of the conspiracy theorist's background beliefs, they cannot possibly serve such a justificatory role. The claim that coherence equals justification can have quite troubling implications when applied to normativity. In reality, abhorrent normative belief systems are often internally contradictory and inconsistent. But this is not logically necessary. Suppose that a person holds an internally consistent Nazi normative worldview. Given the abhorrent nature of this person's beliefs, they cannot establish whether some prospective normative principle is justified. This is not to say that coherence plays no justificatory role – indeed, I will argue shortly that it does – but it cannot be the sole means by which we justify prospective beliefs.

It might be argued that the coherent fairy story is far-fetched. Although there is nothing logically impossible about a person holding a fully coherent set of false beliefs, it seems highly unlikely. Perhaps so, in the case of our descriptive beliefs. But when it comes to our normative beliefs, it might well be that we unwittingly do hold many false beliefs. According to evolutionary debunkers, this situation is not far from the truth. According to Street, evolution has saturated our normative thinking in ways which do not track normative truth. According to Joyce, morality is a myth that we believe just because it provided us with an evolutionary

benefit. In either case, it spells problems for any coherentist methodology. In Street's case, if evolution shaped most of our normative beliefs in directions unrelated to truth, then such off-track background beliefs cannot serve to justify prospective normative principles. Indeed, if a normative principle is shown to cohere with a set of off-track background beliefs that probably counts as evidence that the prospective principle is off-track too. Of course, we established in previous chapters that although many of our normative judgments are caused by evolution, many appear not to be. But this still leaves the issue of how we can determine which ones are and which ones are not. Any suitable normative methodology will need to provide some means by which we can discriminate between or evolutionary and non-evolutionary normative beliefs, otherwise there will be no way of determining which of those beliefs are off-track.

2.3. Foundationalism and Intuitionism

The view I advocate is *Foundherentism*, a third-way epistemological position outlined by Susan Haack, incorporating elements of foundationalism and coherentism (Haack 1993, ch.4). Before I outline a version of foundherentism, I want to say more about foundationalism and especially the role of intuitions, which are central to most versions of this view. According to foundationalism, to avoid an infinite regress we must proceed by establishing some beliefs that are self-standing and do not rely on other beliefs for their justification. These beliefs are recognised by a basic intuition. From these beliefs we can build upwards and account for our wider set of beliefs. Like coherentism, foundationalism can be applied to both normative and descriptive knowledge, though my focus is on normative knowledge. Foundationalism can be either *a posteriori* or *a priori* (perhaps both given that many philosophers, including myself, accept both as legitimate processes by which truths are recognised). An empiricist version of foundationalism will assert that we recognise self-evident beliefs via the senses, rationalist versions will claim that they are recognised intuitively (intermediary views might claim that basic beliefs are supplied both by reason and perception).

For reasons outlined in chapter 1, I argued that realists should accept a non-naturalist view about the nature of normative facts. According to this view, normative concepts like *ought*, *good* and *right* are not the kind of things that we can touch or see, because they are not physical entities and therefore not known via sense experience. Rather, these concepts refer to properties which *supervene* on natural objects. As a non-naturalist who holds a hedonistic view of value, I hold that goodness is a property of the natural phenomenon of pleasure; but that is not to claim that goodness is identical to pleasure. The claim that pleasure is *the good*, is therefore a

complex statement, which denotes natural and non-natural properties. Pleasure is a natural property, but the property ascribed to it, *being good*, is not. This idea of a statement consisting of both natural and non-natural properties might seem peculiar, but we encounter them frequently. Here is an example taken from David Kaspar (2012, pp.135-138): *there are four cows in the field*. *Field* denotes a natural property, as does *cow*. But what about the number four? I can see and touch a cow, but I cannot see and touch the number four. The number four is an abstract, non-natural property that supervenes on natural properties.

How do I know that there are four cows in the field? Either I can detect it via the senses or I might take it on the basis of testimony. Suppose that I leave the field, but I am informed by someone, the local farmer perhaps, that two cows have just escaped from the field. I cannot see the fact that there are now two cows in the field, but I can infer it by conducting a simple mathematical equation ($4-2=2$). The conclusion of this equation is known a priori and does not require any investigation into the nature of the world. Suppose that one of the escaped cows tramples on me. I am in a state of sheer agony. Most of us would accept that being in this state of agony is bad for me. But what is it that we are perceiving in coming to hold this belief? I can perceive the cow trampling on me, I can perceive agony, but I cannot perceive the *fact that agony is bad for me*. Pain's badness is a property I attribute to pain having experienced it.

Consider another example: Singer's *Drowning Child* scenario.¹²⁷ As I walk by a pond, I see a child drowning. The water is shallow enough that it would cause me no danger to wade in and save her, though it might damage my expensive suit. Most people think that in this scenario, the person ought to save the drowning child. Now if this is a fact, as realists might claim, by what means do we come to recognise it? Our sense experience can only take us so far. I perceive the child; I perceive the pond; I am aware that my entering the water is likely to damage my suit; I am aware that should I refuse to enter the pond and retrieve the child, she will likely die. These are all natural facts which I perceive, or infer based on induction, as in the case of my expectation that drowning will result in the child's death, along with any other consequences I foresee. There is no obvious deduction from these facts to the conclusion that I ought to save the child. Neither can it be established inductively, in that our belief would not obviously be strengthened by repeatedly walking past drowning children. At what point do I perceive the fact that I *ought to save the child*? In my view, since *ought* is not a physical property, we never perceive in a sensory way the fact that a person in this scenario ought to save the child

¹²⁷ See Singer (1972).

(especially in cases where, like right now, it is just being considered as part of a thought experiment). Rather, we recognise intuitively that an action which involves a minor sacrifice to save a life has the non-natural property of *being right*, and in so being, has the further property of *being what we ought to do*.

Still, it would be odd to claim that our belief that *agony is bad* and that *saving others at a small cost to ourselves is right* are wholly a priori affairs. How can I know that agony is bad if I have not experienced agony? How can I know I ought to save the child if I have no notion of a child? Though historically some rationalists may have claimed that we can form a comprehensive set of beliefs from the arm-chair, this belief does not characterise contemporary rationalism, whose proponents accept that our knowledge of abstract truths often involves sense experience.¹²⁸ For example, I perhaps need to have the experience of seeing colours such as green and red to know the a priori truth that nothing can be green and red all over. And it is an empirical fact that a gateway into mathematical knowledge often involves demonstrations which appeal to the senses: e.g., putting one object against another and denoting them as “2”.

My rationalist view does not require us to believe that a priori truths (I accept that most of our knowledge is not a priori) are known by reason alone, without any role for the senses. My claim is that some truths cannot be established solely by appeal to the senses, either in the form of basic perception or inductive reasoning. For instance, the fact that nothing can be green and red all over does not deductively follow from my having seen the appearance of green and red things separately; nor is there any obvious inductive argument to that end other than the fact that I have hitherto never encountered a green and red object, but this alone would not account for the apparent impossibility of there being such an object. Nor can this truth be known analytically from the definition of the words red and green. To recognise that something cannot, as a matter of necessity, be green and red all over, requires us to go beyond the empirical premises. It is a truth which we recognise intuitively (if something is green all over, then it is just not possible that it can be red all over). The same is true of normative truths. To recognise that we have reason to save the drowning child of course requires empirical knowledge (about the physical facts, the consequences of alternative actions etc.), but our recognition of that reason is not provided by the empirical facts alone. Intuitionists can accept that the physical facts provide us with the reasons. The phenomenology of agony is what provides us with our

¹²⁸ E.g., Bonjour (1998, pp.9-10).

reason to avoid it: but the natural fact, in this case, the feeling of agony, is not the same thing as *our reason* to avoid agony.

Intuitionism was one of the dominant views of normative ethics in the first half of the 20th century with notable proponents such as Henry Sidgwick, G.E. Moore and W.D. Ross.¹²⁹ Though it retains some notable adherents, its standing within normative ethics is substantially diminished compared with what it was a century ago. Part of this is owed to the dominance of scientific and hard naturalistic views in the latter part of the 20th century. Because I dealt with naturalistic criticisms in chapter one, I will not revisit them here. Intuitionism's decline is also partly the result of criticisms within normative ethics. Some of these criticisms apply only to certain versions of intuitionism and not necessarily to my view. Others are based on misconceptions and mischaracterisations. I will now discuss some of these and attempt to show that, despite its many critics, intuitionism is a viable normative theory.

Intuitionism as Dogmatic

It could be argued that intuitionism is at odds with openminded rational inquiry because it claims that normative beliefs are self-standing and do not derive their justification from other sources. According to this criticism, intuitionists claim that the normative principles in which they believe are known to be self-evidently true as a matter of certainty. But in holding them to be certain, the intuitionist is at odds with the widely held idea that normative beliefs are open to question, which is not in the spirit of rational inquiry (Bonjour 1998, pp.133-137). If the intuitionist really believes that her normative beliefs are indubitable and therefore cannot be challenged, then intuitionism does seem dogmatic. Although some intuitionists may have held this view, it is not a common one among contemporary intuitionists.¹³⁰ Bonjour makes a point of distinguishing his own intuitionist perspective by claiming he does not believe that a priori knowledge is certain or beyond question, nor that such knowledge is infallible (Bonjour 1998, ch.5). Henry Sidgwick made the very same point in *The Methods* (pp.141-142), in which he points out that our confidence in some intuitive belief should be diminished to the extent that we find others do not share that intuition. As I have conceded, our normative judgments are influenced by various distortions and biases and this alone counts against any intuitionist

¹²⁹ See Sidgwick (1907), Moore (2016/1903) and Ross (1930/2002). For discussions of the historical development of intuitionism in this period see Audi (2004, ch.1.) and Kaspar (2012, ch.2). Of the two, Kaspar's discussion focuses more heavily on the criticisms of intuitionism which emerged during this time and contributed to its decline.

¹³⁰ Walter Sinnott-Armstrong (2011, pp.12-13), himself a critic of intuitionism, helpfully distinguishes between two versions of the theory, what he calls *crude* and *sophisticated* intuitionism. Crude intuitionists believe that intuitive normative truths are certain, indubitable and infallible. Sophisticated intuitionists deny this.

claiming that their normative beliefs are indubitable and beyond question. Furthermore, the kind of intuitionism I accept does not restrict intuitions solely to the nature of the proposition being considered, but also to the arguments for and against that proposition, which implies that to establish whether some principle is correct, we need to subject it to maximal scrutiny. I will say more about this at the end of this section.

Intuitionism as Gut-Based

Peter Singer has argued against views that place substantial weight on our common-sense normative intuitions (Singer 2005). Singer has argued, basically on evolutionary grounds, that our common-sense normative beliefs, which are more based on instinct than reflection, are not a good guide toward normative truth. In which case, Singer is a revisionist about the role of normative theory. He does not think that the normative ethicist should seek to explain or justify the pre-reflective normative beliefs we hold. But that we should be prepared to revise and replace them with new ones if that is where the exercise of our reason takes us. As Singer acknowledges, his complaint is not universally applicable to all versions of intuitionism and indeed, in his more recent works, Singer himself has come to accept a version of intuitionism.¹³¹ For the evolutionary reasons outlined, we should approach our common-sense normative beliefs with a healthy dose of scepticism. But not all intuitions are of this instinctive kind. Some of them emanate from a process of intense and critical reflection. Singer endorses Sidgwick's view that the basic axioms of utilitarianism, such as universal benevolence, can be recognised by reason. This intuition conflicts with the more instinctive intuition that we should be partial to people to whom we stand in some special relation. Along with Sidgwick and Singer, the kind of intuitionism I endorse is revisionist. I do not share the view of intuitionist David Kaspar that normative knowledge comes from what people ordinarily think (Kaspar 2012, ch.1). It might be that what we ordinarily think turns out to be correct, but that should not be assumed; the intuitions we have most reason to trust are not necessarily our common-sense intuitions, but those that have been subjected to a high degree of scrutiny and reflection.

Generalism and Particularism

Particularism is a version of intuitionism which is sceptical of the existence of general normative principles (Dancy 2017). A general principle is one which is supposed to operate as

¹³¹ In *The Point of View of the Universe*, de Lazari-Radek and Singer defend Sidgwick's version of intuitionism. The kind of intuitions that should be given weight, on this view, are those of an intellectual nature, as opposed to gut intuitions.

a guide to action across a range of situations. Some generalist theories such as utilitarianism claim that the correct principle applies to all situations. That is, in any possible situation, the correct principle is the one that tells us to maximise happiness. Particularists deny this. They claim that each normative situation supplies us with a particular ought or set of oughts and these do not necessarily carry over into different situations. For instance, it might be that in some situations, it is appropriate to keep a promise even if this reduces overall wellbeing. Whereas in another situation, breaking a promise might be correct. I will not provide a comprehensive argument against particularism here. But suffice it to note that both generalism and particularism are viable options for the intuitionist to take. My view is generalist, insofar as I claim that utilitarianism, a general principle normative theory, can be recognised intuitively. I do think that the analogy between mathematical/logical facts and normative facts, which several intuitionists draw, makes more sense if we accept a generalist account of normative facts. It is not the case that particular situations lend themselves to particular mathematical and logical facts, except in the sense that different objects/forces yield different measurements. Both mathematics and logic provide us with general truths, so if such epistemic domains are akin to normative thought, we should expect normative ethics to provide us with general truths too (otherwise a different analogy is needed).

In claiming that intuitionism is in the business of establishing general truths, it should be kept in mind that my account of intuitionism is unlike that which Sidgwick aptly defines as *perceptual intuitionism* (Sidgwick 1907, bk.1, ch.8). A perceptual intuitionist believes that we intuit normative truths in a manner somewhat akin to sensory perception. If I encounter someone randomly assaulting some innocent person, as an observer of that event I can *just see* the action's wrongness, in a manner akin to seeing that some object is red. I think that critics of intuitionism often have something like this type of view in mind, which I think has contributed to this sense that intuitions must involve a mysterious faculty which allows us to perceive normative properties. On my view, we do not perceive normative properties in any way. To find something intuitively compelling is an act of thought which arises from our general reasoning capabilities and does not presuppose any mysterious cognitive faculty whose sole purpose is to apprehend normative truths.

Intuitionism is Not Tied to any Normative Theory

One potential confusion when considering intuitionism is the idea that it is tied to a particular normative theory. This potential confusion owes itself in part to the way intuitionism has been described by some of its proponents. David Kaspar, for example, claims that intuitionists reject the idea of “supreme principle” theories such as utilitarianism, a theory which he cites as one of intuitionism’s “rivals.”¹³² A supreme principle theory is one which claims that a single principle provides us with the ultimate guide to action, such as the principle of utility. Kaspar claims that “Intuitionism differs from other objectivist moral theories by claiming that there is no need for a supreme principle of morality. We just know what’s right, and thus have no need to deduce, derive, or justify our moral knowledge by means of a supreme principle” (Kaspar 2012, p.4).

Of course, there are intuitionists who reject supreme principle accounts of morality, such as Ross and Kaspar himself. But the claim that intuitionism by its nature rejects supreme principle accounts of morality is plainly false, as is Kaspar’s claim that utilitarianism is necessarily at odds with intuitionism. Indeed, it is puzzling that Kaspar fails to acknowledge this given that he cites Sidgwick, a utilitarian, as a prominent figure in intuitionism. The reason Kaspar is wrong to exclude supreme principle accounts from being intuitionist is that it arbitrarily excludes the view that a supreme principle conception of ethics, like utilitarianism, can be recognised as being true on the basis of intuition(s), a view held by Sidgwick and more recently de Lazari-Radek and Singer. It might well be that this view is incorrect, but it cannot just be ruled out as a conceptual impossibility. Intuitionism takes no particular stance on what the correct normative principles are, only intuitionists themselves can do this. I have expressed support for a broadly hedonistic utilitarian account of ethics, a supreme principle theory, and I think that such a theory can be recognised intuitively (not by a single intuition, but several), though arguing this will go some way beyond the scope of this thesis, as would it be to adequately weigh up the respective merits of supreme principle and pluralistic versions of intuitionism. It is worth noting though, that the supreme principle view is not susceptible to some of the criticisms that have been levelled at pluralistic versions of intuitionism. On Ross’s pluralistic view, there is a range of prima facie duties that generate reasons for action (Ross 1930/2002, ch.2). Unless we have some kind of elaborate lexical ordering of these duties in accordance with their priority, our prima facie duties may clash and it will often be unclear

¹³² Chapter 7 of Kaspar’s book is entitled “Intuitionism’s Rivals” in which he includes utilitarianism.

which of these takes precedence in a given scenario. This problem is avoided by supreme principle views since there is just one principle which guides action.

An Outline of Comprehensive Intuitionism

The intuitionist view I favour is what I call *Comprehensive Intuitionism*. It asserts that our recognition of normative truths does not come wholly from our intuitions concerning the principle itself, but also our broader intuitions concerning arguments for and against that principle. It is often claimed that intuitionists hold normative truths to be self-justifying and non-evidential, insofar as the mere consideration of the content of the principle alone provides us with the knowledge that it is true. This view is at least partially false. Although the content of a principle will need to have intuitive appeal in and of itself, the degree to which it is justified will also depend on the extent to which we find the arguments for and against it intuitively forceful. For example, consider hedonism. I find hedonism intuitively compelling. According to some definitions of intuitionism, if hedonism is true, then its truth is based purely on its intuitive plausibility. The problem with this claim is that it does not allow for our intuitions to be strengthened or weakened in proportion to the arguments for and against hedonism. For instance, many – including myself – find *the experience requirement* intuitively plausible and find that this strengthens their belief in hedonism. If so, then hedonism's truth is not purely self-standing, insofar as it gains further credibility by being supported by other principles and arguments. Likewise, arguments may erode our confidence in hedonism. I am confident enough in the intuitive plausibility of hedonism to accept it, but my confidence is undermined somewhat by *the experience machine objection* (Nozick 1974, p.43), which has some intuitive appeal. This leads me back to the point about certainty. In claiming that hedonism is true, I am claiming that on balance, having tested my intuitions against the principle itself and the arguments for and against, I think that hedonism is true. But I am by no means certain of its truth and my overall confidence in it is reduced by some of my wider intuitions.

This also means that truths known intuitively are not strictly speaking non-evidential. It is important to distinguish between the qualities of being non-evidential and non-inferential. Intuitionist views are non-inferential¹³³ insofar as they claim that intuitive truths are not inferred from prior premises, such as the argument that *if some principle conforms with nature,*

¹³³ Audi seems to suggest that this is historically the hall-mark of intuitionist views (Audi 2004, p.5).

then it is true, egoism conforms with nature, therefore egoism is true. A principle which is non-evidential is one whose justification does not require evidence in favour of that proposition. However, it might well be that a principle is non-inferential, insofar as it is not based on prior premises, but is evidential, insofar as evidence can be appealed to in order to further support or undermine that principle. On the view I accept – and indeed, the very practice of philosophy presupposes – our intuitions count as evidence for or against philosophical hypotheses. If a normative truth is non-evidential, then its truth is based solely on its intuitive credibility. However, if we are consulting our wider intuitions about arguments for and against the principle, then our knowledge of its truth is not strictly speaking non-evidential, because our wider intuitions are being used as evidence. Our initial intuition may not be inferred from evidence, but having formed this intuition, further evidence can be brought to bear to further support or undermine that intuition.

It might be replied that in making this concession, my account is no longer intuitionist. I disagree. It is true that my view is perhaps on the moderate side of intuitionism insofar as I am suggesting that self-evidence alone does not necessarily justify us in accepting some truth and that self-evidence can be defeated by our wider intuitions, or if the apparent self-evidence can be shown to be the result of distortion. But my account is intuitionist insofar as I do allow that self-evidence plays some role in the acquisition of normative knowledge – perhaps the most salient role. The reason I know the sun will rise tomorrow has nothing to do with the inherent plausibility of the proposition – *the sun will rise tomorrow* – but is completely based on the evidence in favour of it. This is not the case when we are considering normative truths. My recognition of hedonism is partly based on its intuitive appeal, but this intuition can be further strengthened or weakened by consulting my wider intuitions. Thus, although I perhaps do not give self-evidence the same degree of salience that previous intuitionists have, my account still qualifies as a version of intuitionism.

It follows that according to comprehensive intuitionism, normative knowledge is neither purely self-standing nor non-evidential. We proceed by considering a normative proposition in a calm and reflective manner. If we find that the principle being considered has intuitive appeal, this provides reason for accepting it (the strength of this reason being proportionate to its intuitive appeal). This is merely one stage in the intuitional process. We then need to consider and test our intuitions against as many arguments and considerations as possible whose conclusions bear upon the truth of that proposition. The outcome (not that the process will ever be truly finished given that new arguments may always be forthcoming) will be that we will either

continue to find the principle intuitively appealing or decide that the arguments against it are sufficient to warrant its rejection (or we may suspend judgment). Based on some of the aforementioned considerations, our determination that it is either correct or incorrect will be a matter of degree; in some cases, we might find the arguments for and against equally compelling and remain undecided or sceptical about the principle. Sometimes we may decide that, on balance, the principle is justified, though significant doubts remain. Other times we may have a high degree of confidence in the truth of the principle.

I want to finish this section by noting that this kind of comprehensive intuitionism is not unique to normative ethics, but characterises philosophy generally. Whether we find, for example, the justified true belief account of knowledge to be correct is basically an intuitive matter. I know of no inductive argument that could settle this question and our observations can only take us so far. Even if I find the theory compelling, the next step will be to consider our intuitions against wider arguments and thought experiments, e.g. Gettier problems and fake barn cases. These two stages of testing our intuitions against the proposition being considered and the arguments and other considerations pertaining to it are not in themselves enough to establish a suitable methodology. We still have not adequately dealt with the problem we started with, which is that some of our intuitions may be distorted by, among other things, the forces of natural selection. I will consider some further components to the methodology to guard against the role of distortive forces within our intuitions. But before doing so, I will settle the question of what, precisely, the structure of this method of justification is.

2.4. Toward a Foundherentist Normative Epistemology?

The account I sketched does not fall neatly within either the coherentist or foundationalist camp. Against coherentism, I have denied that the primary means of justifying principles is to test them against background beliefs. First, we should hypothetically suspend judgment about these beliefs while we consider the content of the prospective normative principle in question, which we should scrutinise on its own terms. Then we proceed to the next stage by testing our intuitions against other relevant considerations. But note that this is not simply a case of testing against our background beliefs. Indeed, whether a background belief can even serve as evidence for the truth of a proposition will be contingent on it passing the first stage described above. Which means that our background beliefs are not automatically given a justificatory role without having been scrutinised themselves. Rather than testing our intuitive principle

against our background beliefs, I think that this stage of the process should be described as the search for reasons to believe and disbelieve the principle we are considering. This is because even if a prospective principle coheres with our existing background beliefs, we should seek to identify new arguments and normative principles which may support or undermine our prospective normative principle. Once a principle has passed these tests (and the ones I will shortly mention) the belief may be classified as a basic normative truth upon which we can build up our picture of normativity. This foundationalist element ensures my account does not fall prey to the infinite regress nor the coherent fairy story objection. There remains the worry that the basic foundational beliefs we find intuitively appealing owe their intuitive force to evolution and other forms of distortion. I will address this problem in the next section.

Against foundationalism, my account leaves an important role for coherence. In the first instance, I denied that any normative belief should be assented to as a matter of certainty, and that, even if we find a principle highly compelling, it is never indubitable. If we find that there are highly intuitive arguments with which that principle conflicts, then this should undermine our confidence in the principle – whether it implies the rejection of the principle will depend on the relative intuitive strength of the objections and the strength of our intuition supporting the principle. It might also be that empirical facts serve an important role here. For instance, I might find a retributive account of punishment highly compelling. But if I come to accept a deterministic account of human action, this empirical fact might serve to undermine my retributivist view, if indeed that view is based on some presupposition about human beings possessing free will.

Moreover, our basic normative beliefs will need to cohere with our wider beliefs about logic. The law of non-contradiction – the belief that nothing can be true and false at the same time – is often touted as one of the most promising candidates of an indubitable self-evident truth which must be presupposed for logical reasoning to take place. If someone asserts, for example, that the principle of utility is both self-evidently true and self-evidently false, then such a combination of beliefs should be dismissed for their failure to cohere with basic logic.¹³⁴ Finally, if we suppose there are several basic normative truths, prospective normative truths will need to be evaluated by their coherence with these truths. For example, suppose I hold the principle of utility to be highly intuitively compelling. Suppose I also come to consider some Aristotelian account of morality, which I also come to find highly compelling. Both of these

¹³⁴ If some principle or theory seems to conflict with the law of non-contradiction, it might not be that we have to reject it. It might be that it can be revised in such a way as to not be contradictory.

views cannot be simultaneously true.¹³⁵ Since the Aristotelian account conflicts with one of my prior basic normative beliefs, I cannot, on pain of incoherence, assent to the Aristotelian view unless I decide, upon reflection, to reject the principle of utility. This leaves a role for a kind of reflective equilibrium in cases where new beliefs conflict with our basic beliefs. Basic beliefs, on my account, are not taken to be indubitable, so there is no reason to think they could not be revised if we come to find other principles more intuitively compelling.

Unlike coherentism and foundationalism, the founderentist approach affords both coherence and self-evidence a justificatory role and in so doing, avoids the excesses and problems of either theory. Since my account does provide a role for both coherence and self-evidence, it appears to follow a founderentist structure. It might be argued that in this way my view is not parsimonious. Though they have faults, at least both foundationalism and coherentism offer a lean explanation of what serves to justify our beliefs. However, the simplest explanation only serves as the best one if it provides an adequate account of what it seeks to explain. As I see it, there are good reasons to think that both coherence and self-evidence play an indispensable role when it comes to epistemic justification. Without any role for self-evidence, there is the issue of an infinite regress. But self-evidence alone seems unduly dogmatic. Surely, if even a highly intuitive belief fails to cohere with the conclusion of some highly compelling arguments against it, then this lack of coherence should at least erode our confidence in the principle. Moreover, the mere acceptance of the law of non-contradiction as having a bearing on what normative beliefs can be justified is already a concession to some kind of coherence. For in accepting this principle, we would need to claim that our basic normative beliefs at least cohere with the belief that no proposition can be both true and false.

3. Dealing with Distortion

3.1. Sinnott-Armstrong's Empirical Case Against Intuitionism

On my account, normative truths are recognised on the basis of intuition – that includes our intuition about the principle itself but also wider arguments and considerations. This raises a pressing question: why should we trust our intuitions? Walter Sinnott-Armstrong argues that empirical evidence can serve to undermine intuitionism (Sinnott-Armstrong 2011). He

¹³⁵ Though it might be possible that elements of both of these views can be simultaneously true if there is no contradiction involved. But on ordinary understandings of these views, they both cannot be true in their entirety because that would involve some contradictions.

highlights evidence showing that our normative intuitions are often flawed and can be influenced by distortive environmental factors. He notes, for example, how the wording of a normative claim can influence people's response to it - e.g., whether some act is phrased as 'saved two of the five' rather than 'not saved three of the five' (ibid, p.16). Sinnott-Armstrong also makes a similar claim to Haidt that emotions such as disgust play a strong role in influencing our normative judgments. And even sleep, or more precisely being deprived of sleep, can influence the normative judgments we make (ibid).

The problem with this argument is that it perhaps explains too much. It is true that our intuitions can be distorted in the aforementioned ways, but then so too can our empirical beliefs. Some people suffer hallucinations; hypnotism might cause someone to believe they have seen something they have not; people can be very bad at estimating risk; some people take perfectly explicable natural events to be a sign of the supernatural. If the mere fact that people can be distorted in their beliefs counts as evidence against them, we should be sceptical about our empirical knowledge. The other issue is that in some of the cases Sinnott-Armstrong mentions, the person is forming intuitions in far from ideal conditions. The point about sleep-levels affecting intuitions seems especially weak. I suspect that depriving students of sleep the night before a science examination is likely to affect their answers and performance. I have already declared that my intuitionist account is rationalistic. So I agree with the contention that intuitions based on gut-based responses to emotions like disgust or hatred should at best be treated with a healthy dose of scepticism. Intuitionists need not claim that all our intuitions are appropriate guides to truth, only those formed in suitable epistemic conditions. It is hard to specify exactly what these conditions are. But we can specify conditions which are not ideal for intellectual reflection, such as being sleep deprived, highly emotional, drunk or unduly deferent to authority (a list by no means exhaustive).

However, it might be objected that this response presupposes that we are able to tell when an intuition is unreliable or influenced by some distortive process. Often it will be difficult to detect distortion, but I do not believe that we should infer that there is no point in trying. Even if we cannot detect the presence of all traces of distortion, it does not follow that we should completely distrust our intuitions. I suspect we will never fully remove all traces of distortion from our sense experience but few argue that we should therefore wholly distrust our senses. However, given that our intuitions can be distorted, an epistemically responsible methodology requires that any prospective intuition is scrutinised for the presence of distortion. For instance, we might consider whether our intuition is motivated by gut-based instinct. If the person

forming the intuition is sleep deprived, we might ask them to reconsider when they are more rested. If someone's intuition changes depending on the wording of some principle, we have reason to suspect that the person is somewhat confused; if they really accepted the principle, then it should make no difference how it is worded.

The rise of the so-called "experimental philosophy" has also brought the use of intuitions in ethics and, indeed, philosophy in general, into question (see Knobe and Nichols 2017). This relates to two different questions; the first, whether intuitions are indeed used by philosophers in their search for truth, the other, as to whether it is appropriate to use intuitions as a method for acquiring knowledge. As to the first, some philosophers deny that intuitions play the important role in philosophy often attributed to them. These so-called "intuition deniers" claim that *arguments* rather than intuitions are the integral component in the philosophical method (Cappelen 2012). Now it is clearly true that arguments play an important role in philosophy; indeed, without the use of rational argument, the practice of philosophy would be all but unrecognisable. But that is besides the point. The intuitionist grants that argument is integral to philosophy, but notes that whether one ultimately comes to recognise a philosophical¹³⁶ argument – e.g. the Gettier argument against the Justified True Belief account of knowledge – as being true requires an appeal to our intuitions.

The second criticism is that even if intuitions play a prominent role in philosophy, the importance they are assigned is regrettable. Some argue that instead of appealing to our intuitions about concepts, we should instead observe and inspect the natural manifestation of our concepts. Take for example knowledge. To learn more about the nature of knowledge, we should not consult our intuitions, we should instead investigate real examples of acquired knowledge. Once again, this criticism gets something right insofar as intuitions are not the be all and end all; the kind of investigation proposed here may be of value insofar as it sheds further light on the nature of knowledge. But it does not thereby follow that intuitions are not important. For instance, before we even get to the process of investigating natural examples of the concept in question, in this case knowledge, we need to decide what things qualify as genuine examples of knowledge and this will require the use of intuitions (Goldman 2015).

3.2. Intuitions and Evolution

¹³⁶ If we are considering an inductive argument, intuition may be less prominent or even non-existent. But for reasons outlined throughout this thesis, I claim that philosophy and ethics are not primarily empirical and conclusions in these domains cannot be established by induction. Hence the reason intuitions are important.

The primary theme of this thesis has been evolution and the potentially distortive role it may have on our normative thinking. According to several of the views I have considered, our normative intuitions are strongly influenced by evolutionary forces - forces which we have good reason to believe are not, for the most part, directed toward normative truth. This suggests that the role of evolution on our intuitions is broadly distorting, though I did note a potential counterexample to this claim in the last chapter. As with the aforementioned distortive influences, the intuitionist will need to provide some account of how we can distinguish between those intuitions that are to be trusted and those which are to be distrusted.

To start this task, it is useful to think about which of our normative beliefs are most likely to be influenced by evolution. According to Haidt's widely endorsed account, evolution shapes our normative beliefs via the emotions. Naturally, this seems to imply that a good starting point is to identify which of our normative beliefs are primarily emotion-based. However, this might not always be easy in practice. Many of us take pride in forming beliefs rationally, which means that even if some belief is caused by emotion, we might not want to admit it to ourselves, or indeed, for reputational purposes, we might not want to admit it to others. An additional issue is that even if emotion is the cause of our normative beliefs, it does not follow that every time we think about the belief in question there is an emotion present. Nor does it follow that we always feel emotion when we express that belief to others. For example, I might hold that breaking promises is bad, primarily due to the anger I felt when some friend broke a promise they made to me. But even though anger was the initial cause, I might no longer feel anger when I warn others of the wrongness of breaking promises. If the initial emotional cause occurred a long time ago, I might now have forgotten that my belief was caused by emotion. Though I think we may sometimes be able to identify emotion's having played a role in some of our normative beliefs, it is not always clear whether it has played the primary role, or indeed, whether the emotion is secondary to the primary cause, as indeed I think it is in the case of the belief that pain is bad, whose primary cause is our aversion to the feeling of pain. Thus, the presence of emotion is at best an imperfect guide toward identifying evolutionary distortion.

A better candidate, perhaps, is the method of seeking out beliefs caused by biases that are likely evolutionary in origin. This is basically to appeal to evolutionary psychology. The evolutionary psychologist claims that some aspects of human psychology can be explained by natural selection and in substantiating these claims, they point to how certain aspects of our psychology would have promoted the survival and reproduction of our ancestors. It seems to follow that according to this logic, we are best able to identify evolution's influence on our normative

beliefs by identifying those which would have provided our ancestors with an evolutionary advantage. I have already discussed some viable candidates for this kind of explanation, such as kin partiality, the in-group bias and belief in retributivism. Moreover, since a good way of telling whether some trait is evolved, as opposed to cultural, is to determine whether it is universal, an obvious candidate for an evolved normative belief will therefore presumably be one that is not specific to any particular culture. Now a caveat is needed here. If we find that some belief is caused by evolution, it does not automatically follow that we should reject it. There are two reasons for this. One is that some normative beliefs are not caused in a way that undermines their justification (I gave an argument in the previous chapter that hedonism falls into this category). Better, I think, to claim that an evolutionary explanation of a principle counts as a reason for scepticism towards it, which puts the onus on the proponent of that belief to explain why, in this case, evolution's role does not serve to undermine it.

The second reason an evolutionary explanation does not automatically imply the rejection of the principle relates to the intrinsic/instrumental distinction. I have claimed that evolution is not a good guide toward normative truth. But how we best achieve the demands of a normative truth might be dependent on the type of creatures we are. To illustrate, let's suppose utilitarianism is the correct ethical theory; in which case, we ought to maximise happiness in any given situation (or, if we are rule utilitarians, to adhere to some set of rules which does). Given the way we evolved, forming close relationships with others makes an enormous contribution to human welfare. But these relationships would not be possible if we did not show partiality to close ones.¹³⁷ Given this, the utilitarian probably should not endorse a general abandonment of kin partiality.¹³⁸ But not all evolved normative beliefs are going to be instrumentally valuable in this way. Though I do not have the empirical evidence at hand to fully substantiate this, I suspect that the in-group bias has served to diminish welfare, in some cases enormously. Thus, if we find that a principle does have an evolutionary explanation, we need to work out whether that principle is intrinsically true, or true only insofar as it is instrumental to the realisation of an intrinsically true principle. If the latter, attention will necessarily turn to the intrinsic principle in question.

¹³⁷ It is for similar such reasons that Sidgwick urges the utilitarian against seeking to demolish the widespread belief in common-sense moral principles. To do so would likely cause more harm than good. This is discussed most thoroughly in bk.4, ch.5 of *The Methods*.

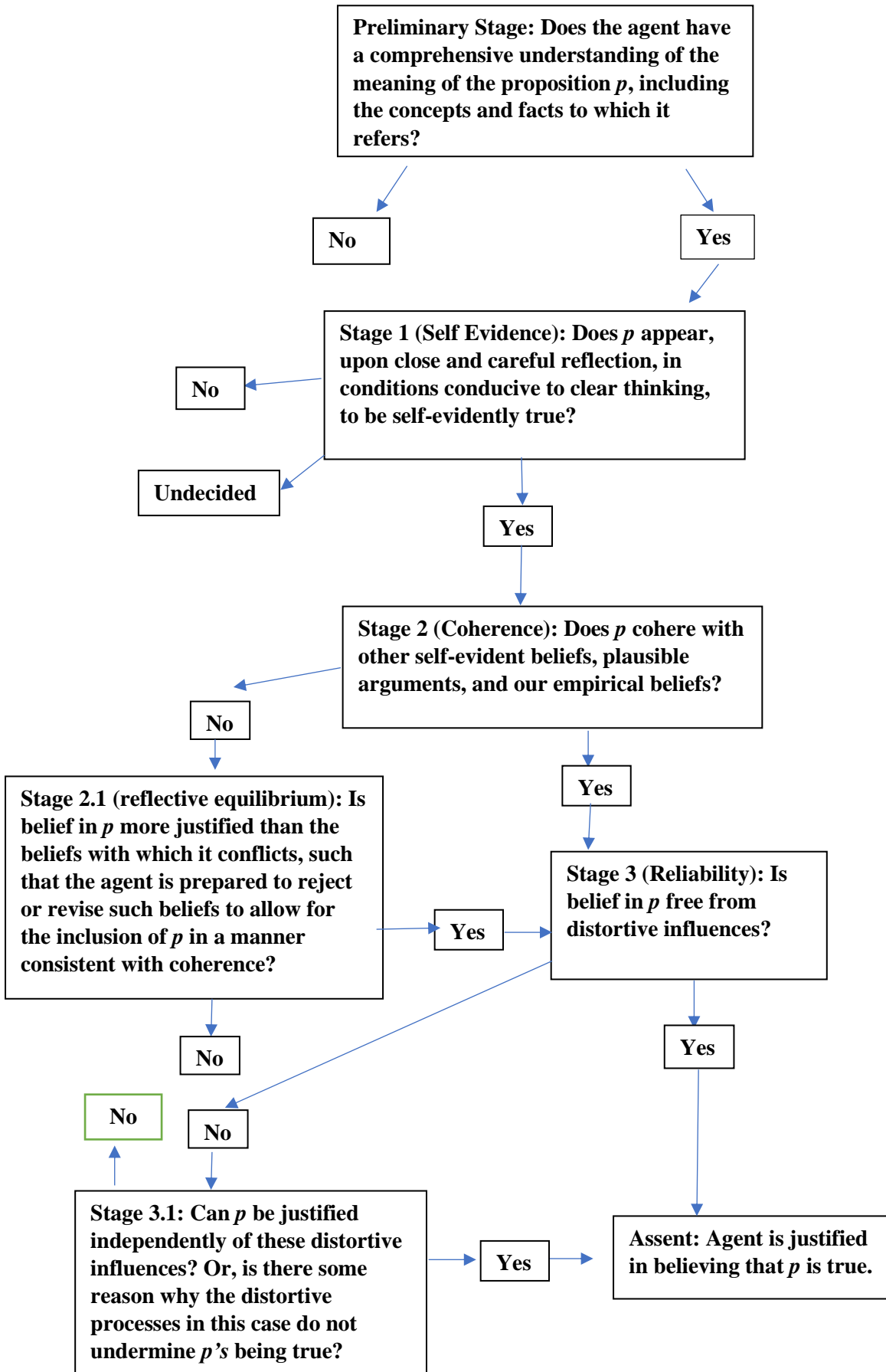
¹³⁸ Though she might claim that it should be somewhat tempered. That we should not prioritise kin over strangers as much as we currently do.

Before I outline a method for dealing with evolutionary and other distortive influences, I want to consider one final proposal for identifying distorted normative beliefs. If evolved normative beliefs are likely to be of a more instinctive nature, it does not seem unreasonable to think that our common-sense normative beliefs are the most likely candidates for an evolutionary explanation.¹³⁹ Common-sense is generally taken to be what we ordinarily think. More revisionist accounts of ethics claim that we should not necessarily defer to our common-sense normative beliefs, but should be prepared to revise or replace them depending on where reason takes us. It does not automatically follow that all common-sense normative views must have evolved; some of what we take to be common-sense might be more culturally conditioned. But at the very least, I think we have grounds for thinking common-sense normative beliefs are more likely to be candidates for evolutionary explanations than more revisionary normative beliefs.

3.3. An Outline of a Normative Methodology

A normative methodology aims to provide some kind of procedure by which we can determine what normative beliefs are justified. Such methodologies are not perfect, in that they will never guarantee that a principle is true. What they do is provide a set of criteria which ensures that our beliefs are being formed in an epistemically responsible way and that possible sources of distortion and confusion are kept at bay. On the next page I provide a diagram illustrating the basic structure of the methodology I advocate, which is followed by a more in-depth discussion of the various stages involved.

¹³⁹ This is not to claim that common-sense beliefs are, by their very nature, products of evolution. Some common-sense beliefs are cultural. To make the distinction between cultural and evolved common sense beliefs, we might identify whether the belief is only common-sense in some specific cultures. If some belief is universally held to be common-sensical, it is a more viable candidate for an evolutionary explanation.



The Preliminary Stage

To know whether some proposition is self-evident it is necessary to understand the concepts involved, whether these be purely abstract or empirical concepts. For example, it is impossible to know that something cannot be red or green all over without understanding the meaning of the words red and green. It might be argued that as a rationalist and a non-naturalist, this kind of understanding can be achieved from the armchair. This is plainly false for reasons that have been outlined. It would be impossible to know that we have an obligation to save a drowning child from a pond without understanding the natural facts been referred to in this sentence. An agent would also need a wider understanding of the consequences which would likely follow such an action; the child dying, the family's grief and so forth. These natural facts can only be known empirically. Sense experience is also important when it comes to evaluating whether mental states are good or bad for us. It is impossible to know whether pain is bad without having the experience of pain.

Stage 1 (Self Evidence): Does p appear, upon close and careful reflection, in conditions conducive to clear thinking, to be self-evidently true?

For the purposes of establishing whether some principle is self-evidently compelling we must try, so far as it is possible, to suspend whatever other beliefs we have which may cause us to be prejudiced for or against the belief in question. For example, to establish whether utilitarian impartiality has intuitive appeal, I must temporarily try to suspend for the purposes of this investigation, my belief in agent-relative partiality - and visa versa if I am a utilitarian attempting to consider whether agent-relative partiality is justified. Why do this? As I have suggested, the first part of my account is foundationalist in nature. Background beliefs only play a role once they themselves have passed the test of self-evidence. Moreover, even if I have, for instance, previously found kin partiality to be self-evident, if I allow this belief to unduly influence me at the first stage of establishing utilitarian impartiality's self-evidence, my judgment will arguably be clouded and the result thereby tainted. There will be plenty of time in stage two for resolving any tensions concerning coherence.

The term "close and careful reflection" stresses the need for patience and epistemic caution. We should not be forming our belief on the basis of gut instinct. Ideally, this process should be repeated several times. Indeed, it is one worth returning to again and again even after a belief has passed all the other stages. It would be irresponsible, for instance, to believe some proposition just because you found it compelling thirty years ago. An epistemically responsible

agent continually re-evaluates and questions their beliefs, including those they hold most dearly. It is not possible to give a comprehensive account of the appropriate conditions under which this investigation should be conducted. Presumably they would be the same kind of conditions which are conducive to basically any intellectual activity. A non-exhaustive list would include:

- Being free from distraction
- Not under the influence of any mind-altering substances
- Impartial
- A state of calm in which emotions are kept in check

I have left open the possibility that the agent is undecided about whether the proposition is true. If this is the outcome, the agent should clearly not progress to the next stage. But this does not mean that the agent should not reconsider whether the principle is true at some future time. The same also applies if the agent has decided that on this occasion, the principle does not seem self-evident. The first stage just asserts that finding a proposition self-evident under these conditions is a sufficient condition for progressing to stage two.

Stage 2 (Coherence): Does p cohere with other self-evident beliefs, plausible arguments, and our empirical beliefs?

Stage 2 is the coherentist stage. Having established that the proposition in question is self-evident, we now need to determine whether there are other considerations, themselves justified, which tend to undermine the proposition in question. This will generally not be an all or nothing affair. We might find that on balance, though some intuitively plausible considerations tend against the proposition we are considering, we still find that proposition to be compelling. For instance, though I find that Nozick's experience machine thought experiment has intuitive appeal, I do not find it strong enough to overturn my belief in hedonism – though I am perhaps less confident in my hedonism than I would have been had I never encountered that argument. Thus, the coherence stage not only serves to establish whether some proposition is or is not justified, but also is instrumental in establishing the degree of confidence we have in it.

I mentioned that for a lack of coherence to be salient, it must be because the belief that conflicts with the proposition we are considering is itself justified. Before I conclude that hedonism's conflict with the experience machine objection is salient, I must first determine whether the experience machine objection passes the first test. Suppose I conclude, on the basis of reflecting

on that objection, that it has intuitive appeal. In these circumstances, I cannot, on pain of incoherence, assent to both the conclusion of the experience machine objection and the truth of hedonism. It is at this stage that the method of reflective equilibrium is incorporated into the methodology. I will need to decide which of these two propositions is most justified. In doing so, I will consider wider arguments against these beliefs, e.g., the asymmetry of our response to the prospect of leaving an experience machine and the idea that our responses to the experience machine are informed by a status quo bias. Basically, it is in reflective equilibrium that the traditional work of normative ethics will be done, where we try to resolve conflicts between plausible normative positions using rational argument.

To summarise stage two. We are determining whether there are considerations which count against the principle we have previously found to be self-evident. If we find that, having considered as many objections as is reasonably possible, none count against the principle we are considering, we may proceed to stage 3 (thus by-passing 2.1.). If as is more likely, we find that there are plausible arguments and considerations with which our principle fails to cohere, then we must try to attempt to resolve these conflicts. In some cases, perhaps we find that no resolution is available, a conclusion which Sidgwick came to concerning the principles of egoism and universal benevolence – this may imply a state of scepticism or, if it is possible, require some means of including conflicting principles by lexical ordering. We might find that the wider arguments strongly outweigh the initial intuitive appeal of the principle that passed stage 1, in which case, we should not proceed to stage 3. Finally, if we find that the principle can be adequately defended against plausible objections, we are justified in proceeding to the next stage – though it might be that by this point, assuming some of the counterarguments are strong, we are somewhat less confident in the principle.

Stage 3 (Reliability): Is belief in p free from distortive influences?

By this point we have established the principle being considered is self-evident and that it either does not face any compelling objections, or if it does, can adequately be defended against them. However, it might be that we still have reason to be sceptical about this principle if it can be shown that its intuitive appeal owes itself to some kind of distortive influence. This stage of the method is reliabilist in nature (Goldman 2009). Ideally, we would show the process by which we came to find the principle to be self-evident is reliable. In practice it might be too difficult to locate the exact *source* of our belief. Perhaps a more achievable aim will be to establish whether there is evidence of some distortive influence at work. If there is no such

evidence, then we are justified in assenting to the principle, assuming it has passed the other stages. This does not mean that we can ever know for certain whether our belief is distorted. But the mere possibility cannot be enough to warrant scepticism unless we are prepared to be perpetually sceptical about all our beliefs. A proportionate approach would be to assent to the belief in the absence of evidence for a distortive influence whilst accepting that at some future stage, such evidence may become available, meaning stage 3 will need to be revisited.

By what process can we know whether the principle is the result of distortion? To this there is no straightforward answer (though stage 1 requires that the belief be formed in the absence of some obvious sources of distortion, such as drunkenness and inflamed passion) and the means of detection may vary depending on the type of distortion being investigated. If we are trying to determine whether the intuitive appeal of the principle is caused by natural selection, we might seek to determine whether there is a plausible evolutionary explanation of the belief which shows how it might have benefitted our ancestors. If no such evidence is available, we may conclude that there is no reason to believe the intuitive appeal of the principle is caused by evolution (note that this is not the same as denying such an explanation altogether). Even if such an explanation is identified, further evidence might be brought to bear. If some trait is evolved as opposed to cultural, then it is likely to exist across cultures. In which case, as well as having a good evolutionary explanation for some normative belief, a potential debunker may also need to show that it is cross-cultural, or if it is not, give a good explanation for why it is not. This is just one example. Beliefs can be distorted by a wide range of causes. A complete methodology will attempt to fill in all these gaps. But I have insufficient space to do this here and I suspect this task will be an on-going one as we discover more about the processes involved in our normative judgments. I am simply stressing the need for a stage in our methodology which attempts to identify the distortive influences.

It will be noted that on the method proposed above, the mere detection of a distortive influence is not a sufficient condition for our rejection of the principle. I leave open the possibility, in stage 3.1, that in spite of the evidence for the role of a distortive process, we can appeal to external reasons which justify the principle, or attempt to explain away the apparent distortive influence. This, I am afraid, is the vaguest aspect of the method I have proposed, which is undoubtedly a defect. In chapter 4, I tried to provide an argument of this kind which I believe shows a fairly unique way in which a generally distortive influence (evolution) does not serve to undermine a belief (hedonism) in which it has played a causal role. The explanation is quite technical and would not necessarily cross over to other forms of distortion. For the sake of

variety, here is another possible way in which some form of distortion might be explained away without damaging the credibility of the principle being considered. Suppose that after reflecting on the cause of some strongly held belief of mine, I remember that I first encountered it after hearing it expressed by someone who has a disastrous track record of forming true beliefs. In which case, the cause of my belief is an unreliable source. But it seems to me that this should have no bearing on whether I am justified in believing the principle. True, I might have more reason to believe it if I was informed of it by a more reputable authority. But this explanation does not seem to undermine the fact *that I find the principle intuitive*, so long as my attraction to it is based upon my having reflected on the content of the principle, rather than gullibly accepting the testimony of an unreliable source.

Ultimately, I do not want to pre-empt all the possible ways in which some apparent distortive influence might be explained away. To claim that only such and such explanations are available would be too restrictive and therefore counterproductive. My position is that if there is evidence that one's intuitively held belief is caused by distortion of some kind, the onus is then on them to provide a satisfactory account for why in this case, the evidence does not undermine the principle. That I have not provided precise criteria that such a defence must meet means that there is a certain openness to my methodology, which means it is not a 'closed system.' In determining whether a defence of this kind is viable, the final court of appeal may need to be our intuitions.

Conclusion

The first part of this chapter consisted of my response to the content-neutral evolutionary debunking argument. Rather than provide a decisive argument against this view, I instead provided a rival genealogical account of normative thinking which provides a fuller explanation of its content. The evolutionary genealogy of our normative thinking is best applied to the aspect of our normative thinking which involves following and adhering to rules. Evolution also helps to explain our pre-reflective instinctive normative beliefs. But this type of explanation falls short when it comes to explaining the more questioning aspect of our normative thinking, which does not involve blindly accepting normative truths but a search for the correct truths and a critical attitude toward the received wisdom. According to my account, this aspect of our normative thinking is best explained by our ability to engage in critical reasoning of a kind exemplified by the activity of philosophy. This dualistic genealogy, which

gives due weight to evolution, provides a more complete account of the various dimensions of our normative thinking.

In the second half, I discussed questions pertaining to epistemology and methodology. The first asks what the structure of our normative knowledge is. The second, by what method we can go about justifying normative principles. I argued that both foundationalism and coherentism have their advantages and defects. Appealing to coherence alone is insufficient because we need to first establish whether we can trust our background beliefs; a point made more salient by the possible distortive role of natural selection. But coherence clearly must play some role: if some intuitive principle fails to cohere with intuitive counterarguments, this is clearly an issue for the principle. Given these considerations, I suggested that the only viable account of the structure of normative knowledge will be one which gives some weight to both self-evidence and coherence. In this respect, the account I proposed is foundherentist.

In section three, I proposed a viable methodology for justifying normative principles. I set out three stages, preceded by a preliminary stage. In the preliminary stage we must have a firm understanding of the proposition being considered. Then we proceed to stage one where we assess the belief for self-evidence. When this foundationalist stage has been completed, we move to the coherentist stage, testing the principle against an array of compelling arguments. If we find that the principle can be defended against any undermining considerations, we proceed to the reliabilist stage. We seek to establish whether there is evidence of distortive processes being responsible for our intuitive belief in the principle being considered. If there are none, then we are justified in accepting the principle – though we might find reasons to reject it in the future. If we find evidence of distortion, we then need to ascertain whether there is a compelling defence of that principle available, notwithstanding the apparent causal role of distortive processes. If this can be done, as indeed I believe I did in chapter 4 in my defence of a version of hedonism, then the agent is justified in accepting that principle as being true. If it cannot, the principle should be rejected.

Concluding Remarks

This thesis has defended non-naturalistic realism against a range of arguments, but my primary focus has been defending it against EDA. In the first chapter, I provided a statement of non-naturalistic realism. This is the view that normative facts exist, that their truth does not depend in any way on the evaluative judgments of agents, that such truths are non-natural abstract truths as opposed to natural facts, and that they can be recognised a priori. I provided some reasons for why this account of realism has more going for it than its naturalistic counterpart, which considers normative facts as a species of the natural, whose truths are known empirically. I stress in the final chapter that empirical knowledge clearly plays some role in our acquisition of normative knowledge. One could not recognise the wrongness of murder without understanding the natural facts which constitute an act of murder, neither can one recognise agony's disvalue without experiencing agony. However, sense experience alone can only take us so far. Without violating Hume's is/ought distinction, there is no logical progression from descriptive premises about what is the case to a normative conclusion about what ought to be. To bridge the gap, we must appeal to rational insight in the form of intuition.

Having defended this view against a range of well-known objections in chapter one, I turned my focus in chapter two toward EDA. These are arguments which, in their metaethical form, seek to undermine normative realism by highlighting the distortive role evolution has played in shaping human normativity. There I drew a distinction, to my mind original, between *content-neutral* and *content-partial* EDA. Sharon Street's influential *Darwinian Dilemma* argument against realism is an example of the latter. Street sought to undermine realism from the ground up, by highlighting the role evolution has played in shaping our first-order normative beliefs. Evolution is not in the business of tracking stance-independent truths, according to Street, because even if they exist, they would not contribute to the survival and reproduction. But evolution is very much in the business of shaping our normative beliefs, but it does so in whatever way contributes to evolutionary fitness. Assuming this does not correlate with normative truth, we might suppose that evolution's influence has been normatively off-track. And in so being, only an unlikely coincidence could have resulted in those beliefs turning out to be true. Joyce's EDA differs in terms of its premises and conclusion. Joyce's premises do not include Street's claim that evolution shaped the content of our first-order normative beliefs. Rather, Joyce remains neutral about whether evolution has shaped individual first-order normative beliefs. His claim is that evolution is responsible for our general tendency to think

normatively and the belief that there are categorical normative reasons for acting. According to Joyce, this belief is a useful fiction we believe only because it motivated our ancestors to transcend self-interest and act in ways that contribute to the good of the group. His conclusion differs from Street's in that whereas Street is just objecting to the realist's stance-independent understanding of normative truth, Joyce thinks the evolutionary origins of normative thinking undermines both anti-realist and realist understandings of normative truth.

Because these arguments differ in content, they require different responses. My responses to Sharon Street's content-partial EDA take place in chapters three and four. In chapter three, I respond to Street's claim that evolution overwhelmingly shaped our normative beliefs. First, unless evolution shaped all our normative beliefs – a belief Street does not appear to hold – then even if its influence is distortive, it would not undermine all our normative beliefs. In particular, it would not serve to undermine those beliefs – however small a number – which it did not shape. Moreover, there are beliefs which appear to fit this description, insofar as they do not lend themselves to an evolutionary explanation. Some of those I mentioned include utilitarian impartiality, rejection of the acts and omissions distinction, and commitment to value maximisation. Even if just these beliefs are excluded from evolutionary explanations – I suggested there are others – we have already gone considerably toward immunising utilitarianism from EDA.

My revised version of the argument from disagreement brought further limitations on the claim that evolution overwhelmingly shaped our normative beliefs. Humans share roughly the same evolutionary heritage. Our Stone Age ancestors were all responding to basically the same environmental conditions and according to evolutionary psychologists, these stone age pressures hold considerable explanatory power when it comes to human psychology. But if our normative views were shaped by basically the same evolutionary pressures, then is it not a mystery how humans came to hold such radically different views? If normative beliefs were overwhelmingly caused by evolution, we should expect a high degree of uniformity. That this does not obtain highlights the need for a more pluralistic normative genealogy which attributes our normative beliefs to various sources. I accept that evolution played a strong role in shaping our normative beliefs, but it was by no means the only factor. A claim I further substantiate in the final chapter.

In chapter four I respond to another of Street's claims. Namely, that if evolution has shaped a normative belief, that belief is likely to be off-track. This claim is, broadly speaking, a plausible

one. But there are counterexamples. One of which being the hedonic belief that being in a state of aversion toward one's pain is bad. The asymmetry arises out of the fact that evolution is not uniform in the way that it shapes our normative beliefs. According to Jonathan Haidt's social intuitionist account of normative judgment, one that has been highly influential among evolutionary psychologists and metaethicists, evolution shapes normative beliefs via intermediary emotional states. In which case, one can make the case that evolution *shapes*, *tricks* or *misleads* (depending on one's view about the role of emotion in ethics) us into holding particular normative views. Though I accept that this model can be applied to many of our normative beliefs, it cannot as easily be applied to hedonic beliefs. Our belief that *being in a state of aversion toward pain is bad* is based primarily on our experience of what it is like to have that sensation. Emotions may play some role, particularly in the aftermath of that experience, or our anticipation of it. But it is conceptually possible to disentangle our aversion to pain and our emotional experience of it. Insofar as it is our direct acquaintance with aversion to pain which leads us to conclude that it is bad, rather than emotion, this belief is more direct and therefore more epistemically secure. This leads me to the convenient albeit justified conclusion that hedonism, though in some sense caused by evolution, is not undermined by its evolutionary origins. Combine this with my defence of utilitarian principles in the preceding chapter and we find that realists who accept a hedonistic utilitarian view of ethics escape Street's content partial EDA.

In the final chapter, I began by considering Joyce's content-neutral EDA. Joyce is neutral about whether evolution shaped the content of our first order normative beliefs, though he thinks the very fact that we hold normative beliefs in the first place can be explained by evolution. To some degree, Joyce is correct. We all hold instinctive normative beliefs. Such beliefs broadly characterise common-sense morality. These beliefs are often based on our emotional responses, rather than reflection. If you fail to keep a promise you made to me, I am likely to get angry and accuse you of wrongdoing. Though I could probably rationalise my disapproval of your failure to keep the promise, my initial disapproval is not based on rational argument, but a natural pre-reflective response. But we can also form normative beliefs in a more critical and reflective way and sometimes these beliefs conflict with our instincts. Though I might instinctively disagree with your action, on reflection, I might come to wonder whether there really is an obligation to keep promises, and if so, whether this obligation can be defeated in certain circumstances. Joyce's account, which is based on the role of our moral emotions, provides a compelling explanation of the kind of instinctive, gut-based pre-reflective normative

thinking which characterises common-sense morality. It is less convincing when applied to our more reflective normative beliefs, whose origins lie in our ability to reason – an ability which is more difficult to debunk because it has a track record of tracking truth.

Given that my account concedes that evolution has had considerable influence on our normative judgments, some account needs to be given of how we can know whether some prospective normative principle is undermined by belief in it being the result of distortive influences. There is no straightforward means by which to do this, but in the second half of the final chapter I attempted to provide a viable methodology which assists with this goal. My account is foundherentist in structure, insofar as it takes elements from both coherentist and foundationalist epistemologies. In the first instance, we assess the normative principle for self-evidence. If we deem it self-evident, we progress to the next stage where we determine whether it coheres with other intuitively compelling normative principles and arguments. This stage is the one where the traditional work of normative ethics takes place, where we consider the merits and defects of a range of arguments against the principle we are considering. The third stage of my method is where we aim to identify distortion, including the role of evolution. We employ methods developed by evolutionary psychologists to seek to identify whether some normative belief might have its origins in evolution, because it promoted the survival and reproduction of our ancestors. If no such evolutionary explanation is available, we are justified in assenting to the principle. If such an explanation is available, we need to work out whether some explanation can be given for why, in this instance, evolution's influence does not distort the principle, as I did for hedonism in chapter four. If no such explanation is available, we have strong reason reject the principle.

Despite their enormous influence over the last two decades, EDA do not undermine normative realism. They might undermine certain realist views, for example, those combined with first-order normative beliefs that can be debunked by evolution. But not all realists fall into this category. Indeed, my view does not fall into this category. I hold that hedonistic utilitarianism is true. Most aspects of this theory cannot be explained by evolution. It does not make sense, from a purely evolutionary perspective, to believe that my interests are no more intrinsically important than the interests of strangers, nor that I have an obligation to make enormous personal sacrifices for the wellbeing of strangers. Yet this is precisely why realists who hold such views are not being misled by evolution: evolution would have us think differently. And the aspect of this theory which can be explained, in some respects, by evolution is none the worse for it. Though I would not recognise that hedonism is true had I not evolved to feel pain,

the fact that I can feel it means I have a direct insight into its nature. And it is through this direct insight that I acquire my belief that its bad, not via an intermediary emotional state. I have not earned the right to assert that this view has been proven correct, since I have offered no positive argument for it. But I have earned the right to claim that this view cannot be debunked by evolution. This thesis has engaged with EDA in relation to metaethics, normative ethics, and epistemology. I provided an original way of framing the two key main forms of EDA and in chapters 3 and 4 I provide several novel lines of response to EDA. My final key achievement in this thesis is the normative methodology in chapter 5, which combines insights from three of the most influential epistemological theories (foundationalism, coherentism, reliabilism). My ambition was to outline a pluralistic methodology that avoids the pitfalls these theories may suffer individually and guards against distortive influences such as evolution. My hope is that these insights will contribute to the ongoing debate between realists and anti-realists and in particular, show that despite the force of EDA, normative realism remains a credible metaethical option.

Bibliography

- Archer, Alfred. 2018. "Supererogation." *Philosophy Compass* 13:3. 1-9. DOI: 10.1111/phc3.12476
- Aristotle. 1926/340BC. *Nicomachean Ethics*. Translated by H. Rackham. Cambridge: Harvard University Press.
- Audi, Robert. 1996. "Intuitionism, Pluralism, and the Foundations of Ethics." In *Moral Knowledge? New Readings in Moral Epistemology*. 101-136. Edited by Mark Timmons and Walter Sinnott-Armstrong. Oxford: Oxford University Press.
- Audi, Robert. 2004. *The Good in the Right: A Theory of Intuition and Intrinsic Value*. Princeton: Princeton University Press.
- Ayer, Alfred Jules. 1952. *Language, Truth and Logic*. Second Edition. New York: Dover Publications.
- Bain, David. 2014. "Pains That Don't Hurt." *Australasian Journal of Philosophy* 92:2. 305-320. DOI: 10.1080/00048402.2013.822399.
- Berman, Mitchell N., 2013. "Rehabilitating Retributivism." *Law and Philosophy* 32:1. 83-108. DOI: 10.1007/s10982-012-9146-1.
- Boehm, Christopher. 1999. *Hierarchy in the Forest: The Evolution of Egalitarian Behaviour*. Cambridge: Harvard University Press.
- Bradford, Gwen. 2020. "The Badness of Pain." *Utilitas* 32:2. 236-252. doi:10.1017/S0953820819000475
- Bramble, Ben. 2017. "Evolutionary Debunking Arguments and Our Shared Hatred of Pain." *Journal of Ethics and Social Philosophy* 12:1. 94-101. DOI: 10.26556/jesp.v12i1.214.
- Brink, David O., 2001. "Realism, Naturalism and Moral Semantics." *Social Philosophy and Policy*. 18:2. 154-176. DOI:10.1017/S0265052500002946.
- Bronsteen, John. 2017. "Preferring to Decrease One's Own Wellbeing." *Utilitas* 29:1. 52-64. DOI: 10.1017/S0953820816000212
- Brooke, John Hedley. 2009. "Darwin and Victorian Christianity." In *The Cambridge Companion to Darwin*. 197-218. Edited by Jonathan Hodge. Cambridge: Cambridge University Press.
- Brüllman, Philipp. 2013. "Good (as) Human Beings." In *Aristotelian Ethics in Contemporary Perspective*. 97-113. Edited by Julia Peters. New York: Routledge.
- Bonjour, Laurence. 1998. *A Defence of Pure Reason*. Cambridge: Cambridge University Press.
- Burkhart, Judith M., Rahel K. Brügger and Carel P. van Schaik. 2018. "Evolutionary Origins of Morality: Insights from Non-Human Primates." *Frontiers in Sociology* 3:17. 1-12. DOI: 10.3389/fsoc.2018.00017

- Buss, David M. 2012. *Evolutionary Psychology: The New Science of the Mind*. 4th Edition. Boston: Pearson.
- Cappelen, Herman. 2012. *Philosophy Without Intuitions*. Oxford: Oxford University Press.
- Clarke-Doane, Justin. 2012. "Morality and Mathematics: The Evolutionary Challenge." *Ethics* 122:2. DOI: 10.1086/663231.
- Coliva, Annalisa. 2019. "Disagreeing with Myself: Doxastic Commitments and Intrapersonal Disagreement," *American Philosophical Quarterly* 56, 1:1-13. ISSN: 00030481.
- Copp, David. 2007. *Morality in a Natural World: Selected Essays in Metaethics*. Cambridge: Cambridge University Press.
- Copp, David. 2008. "Darwinian Skepticism About Moral Realism." *Interdisciplinary Core Philosophy* 18. 186-206. ISSN: 15336077.
- Crisp, Roger. 2006. *Reasons and the Good*. Oxford: Oxford University Press.
- Daly, Martin and Margo Wilson. 2005. "Carpe Diem: Adaptation and Devaluing the Future." *The Quarterly Review of Biology* 80:1. 55-60. DOI: 10.1086/431025
- Dancy, Jonathan. 2017. "Moral Particularism." *Stanford Encyclopaedia of Philosophy*. <https://plato.stanford.edu/entries/moral-particularism/>. Date accessed: 27/07/2020.
- Daniels, Norman. 1996. *Justice and Justification: Reflective Equilibrium in Theory and Practice*. Cambridge: Cambridge University Press.
- Darwin, Charles. 1859. *On the Origin of Species by Means of Natural Selection, Or The Preservation of Favoured Races in the Struggle for Life*. London: John Murray.
- Darwin, Charles. 2004/1879. *The Descent of Man, and Selection in Relation to Sex*. London: Penguin.
- Darwin Correspondence Project. "Letter No.2743." Accessed: 11th January 2021. Link: <https://www.darwinproject.ac.uk/letter/DCP-LETT-2743.xml>.
- Dawkins, Richard. 2003. *A Devil's Chaplain: Selected Essays by Richard Dawkins*. London: Weidenfield and Nicholson.
- Dawkins, Richard. 2016/1976. *The Selfish Gene*. 4th Edition. Oxford: Oxford University Press.
- Dennett, Daniel C., 1995. *Darwin's Dangerous Idea*. London: Allen Lane.
- Descartes, Rene. 1641/1993. *Meditations on First Philosophy in which the Existence of God and the Distinction of the Soul from the Body are Demonstrated*. Third Edition. Translated by Donald A. Cress. Cambridge: Hackett.
- Dworkin, Ronald. 2011. *Justice for Hedgehogs*. Cambridge: Harvard University Press.
- Ellemers, Naomi., Jozanneke van Der Toorn, Yavor Paunov and Thed van Leeuwen. 2019. "The Psychology of Morality: A Review and Analysis of Empirical Studies Published From 1940 Through 2017." *Personality and Psychology Review* 23:4. 332-366. DOI: 10.1177/1088868318811759

- Field, Hartry. 2016. *Science Without Numbers: A Defence of Nominalism*. 2nd Edition. Oxford: Oxford University Press.
- Fletcher, Guy. 2019. "Taking Prudence Seriously." In *Oxford Studies in Metaethics: Volume 14*. 70-94. Edited by Russ Shafer-Landau. Oxford: Oxford University Press.
- Foot, Philippa. 1967. "The Problem of Abortion and the Doctrine of Double Effect." *Oxford Review* 5: 5-15.
- Gampel, Eric H., 1996. "A Defence of the Autonomy of Ethics: Why Value is Not Like Water." *Canadian Journal of Philosophy* 26:2. 191-210. DOI: 10.1080/00455091.1996.10717450.
- Geach, Peter. 1965. "Assertion." *Philosophical Studies* 74:4. 449-465. DOI: 10.2307/2183123.
- Gettier, Edmund. 1963. "Is Justified True Belief Knowledge?" *Analysis*. 23:6. 121-123. DOI: 10.1093.
- Gibbard, Allen. 1990. *Wise Choices, Apt Feelings*. Oxford: Clarendon Press.
- Gibbard, Allen. 2017. "Parfit on Normative Concepts and Disagreement." In *Does Anything Really Matter? Essays on Parfit on Objectivity*. 61-80. Edited by Peter Singer. Oxford: Oxford University Press.
- Goldman, Alvin I. 2009. "Internalism, Externalism, and the Architecture of Justification." *The Journal of Philosophy*. 106:6. 309-338. DOI: 10.5840.
- Goldman, Alvin I. 2015. "Review of Hilary Kornblith, *A Naturalistic Epistemology: Selected Papers*." *Notre Dame Philosophical Reviews*.
- Goldstein, Irwin. 1989. "Pleasure and Pain: Unconditional, Intrinsic Values." *Philosophy and Phenomenological Research* 50:2. 255-276. DOI: 10.2307/2107959.
- Godwin, William. 1793/2001. *Enquiry Concerning Political Justice*. Ontario: Batoche Books.
- Greene, Joshua. 2014. *Moral Tribes: Emotion, Reason and the Gap Between Us and Them*. London: Atlantic Books.
- Haack, Susan. 1993. *Evidence and Inquiry: Towards Reconstruction in Epistemology*. Oxford: Blackwell.
- Haack, Susan. 2002. "A Foundherentist Theory of Empirical Justification." In *Epistemology: Contemporary Readings*. Edited by Michael Huemer. 417-431. London: Routledge.
- Haidt, Jonathan. 2001. "The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgment." *Psychological Review* 108:4. 814-834. DOI: 10.1037/0033-295X.108.4.814.
- Hall, Richard J., 1989. "Are Pains Necessarily Unpleasant?" *Philosophy and Phenomenological Research* 49:4. 643-659. DOI: 10.2307/2107852.
- Hamilton, William Donald. 1964. "The Genetic Evolution of Social Behaviour." *Journal of Theoretical Biology* 7:1. 1-16.

- Hare, Richard Mervyn. 1991. "Universal Prescriptivism." In *A Companion to Ethics*. 451-463. Edited by Peter Singer. Oxford: Wiley-Blackwell.
- Huemer, Michael. 2016. "A Liberal Realist Answer to Debunking Sceptics: The Empirical Case for Realism." *Philosophical Studies* 173:7. 1983-2010. DOI: 10.1007/s11098-015-0588-9.
- Hume, David. 1985/1739. *A Treatise of Human Nature*. Edited by Ernest Mossner. London: Penguin.
- Huxley, T.H., 1995. "Evolution and Ethics." In *Issues in Evolutionary Ethics*. 111-150. Edited by Paul Thompson. New York: State University of New York Press.
- Jackson, Frank. 2003. "Cognitivism, A Priori Deduction and Moore." *Ethics*. 113:3. 557-575. DOI: 10.1086/345936.
- Jamieson, Dale. 2002. "Is there Progress in Morality?" *Utilitas* 14:3. 318-338. DOI: 10.1017/S0953820800003630.
- Joyce, Richard. 2001. *The Myth of Morality*. Cambridge: Cambridge University Press.
- Joyce, Richard. 2006. *The Evolution of Morality*. Massachusetts: MIT Press.
- Joyce, Richard. 2013. "Irrealism and the Genealogy of Morals." *Ratio* 26:4. 351-372. DOI: 10.1111/rati.12027.
- Kahane, Guy. 2011. "Evolutionary Debunking Arguments." *Nous* 45:1. 103-125. DOI: 10.1111/j.1468-0068.2010.00770.x
- Kahane, Guy. 2017. "Pain, Experience, and Wellbeing." In *The Routledge Handbook of Philosophy of Wellbeing*. Edited by Guy Fletcher. 209-220. London: Routledge.
- Kaspar, David. 2012. *Intuitionism*. London: Bloomsbury.
- Kennedy, Kelly. 2010. *They Fought for Each Other: The Triumph and Tragedy of the Hardest Hit Unit in Iraq*. New York: St. Martin's Press.
- Kitcher, Philip. 2011. *The Ethical Project*. Harvard: Harvard University Press.
- Klenk, Michael. 2018. "Third Factor Explanations and Disagreement in Metaethics." *Synthese* 197. 427-446. DOI: 10.1007/s11229-018-1875-8.
- Knobe, Joshua and Shaun Nichols. 2017. "Experimental Philosophy." *The Stanford Encyclopaedia of Philosophy*. <https://plato.stanford.edu/cgi-bin/encyclopedia/archinfo.cgi?entry=experimental-philosophy>. Date Accessed: 17/09/2022.
- Korsgaard, Christine. 2006. "Morality and the Distinctiveness of Human Action." In *Primates and Philosophers: How Morality Evolved*. Edited by Stephen Macedo and Josiah Ober. 98-119. Princeton: Princeton University Press.
- Kripke, Saul A., 1980. *Naming and Necessity*. Second Edition. Oxford: Basil Blackwell.
- Kropotkin, Peter. 1902. *Mutual Aid: A Factor in Evolution*. McClure, Philips & Company.

- Locke, John. 2004/1690. *An Essay Concerning Human Understanding*. Edited by Roger Woolhouse. London: Penguin Books.
- Mackie, John Leslie. 1977. *Ethics: Inventing Right and Wrong*. Middlesex: Penguin.
- Manne, Kate. 2017. "Locating Morality: Moral Imperatives as Bodily Imperatives." In *Oxford Studies in Metaethics Volume 12*. Edited by Russ Shafer-Landau. 1-26. Oxford: Oxford University Press.
- Marlowe, Frank W., 2005. "Hunter-gatherers and Human Evolution." *Evolutionary Anthropology* 14. 54-67. DOI: 10.1002/evan.20046.
- Marx, Karl. 1959/1844. *Economic and Philosophic Manuscripts of 1844*. London: Lawrence and Wishart.
- Midgley, Mary. 2014. *The Solitary Self: Darwin and the Selfish Gene*. Oxon: Routledge.
- Mill, John Stuart. 1863/2015. "Utilitarianism." In *On Liberty, Utilitarianism and Other Essays*. Edited by Mark Philp and Frederick Rosen. 115-180. Oxford: Oxford University Press.
- Mill, John Stuart. 1874. "Nature." In *Nature, The Utility of Religion and Theism*. Second Edition. London: Longmans, Green, Reader and Dyer.
- Miller, Alexander. 2013. *Contemporary Metaethics: An Introduction*. Second Edition. Cambridge: Polity Press.
- Moore, G.E., 2016/1903. *Principia Ethica*. Jovian Press.
- Morton, Justin, 2019. "When do Replies to the Evolutionary Debunking Argument Against Moral Realism Beg the Question?" *Australasian Journal of Philosophy* 97:2. 265-280. DOI: 10.1080/00048402.2018.1455718.
- Naess, Arne. "The Shallow and the Deep, long-range ecology movement. A summary." *Inquiry* 16. 1-4 (1973): 95-100.
- Nagel, Thomas. 1974. "What is it Like to be a Bat?" *The Philosophical Review* 83:4. 435-450. DOI: 10.2307/2183914.
- Nagel, Thomas. 1986. *The View from Nowhere*. Oxford: Oxford University Press.
- Nietzsche, Friedrich. 1887/2012. *On the Genealogy of Morals*. Translated by Ian Johnston. Cranston: Penguin Classics.
- Nozick, Robert. 1974. *Anarchy, State and Utopia*. New York: Basic Books.
- Nussbaum, Martha. 2000. *Women and Human Development: The Capabilities Approach*. Cambridge: Cambridge University Press.
- Odenbaugh, Jay. 2017. "Nothing in Ethics Makes Sense Except in the Light of Evolution? Natural Goodness, Normativity, and Naturalism." *Synthese*. 194. 1031-1055. DOI: 10.1007/s11229-015-0675-7.
- Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press.
- Parfit, Derek. 2011a. *On What Matters: Volume 1*. Oxford: Oxford University Press.

- Parfit, Derek. 2011b. *On What Matters: Volume 2*. Oxford: Oxford University Press.
- Parfit, Derek. 2017. *On What Matters: Volume 3*. Oxford: Oxford University Press.
- Perry, R. B., 1970/1926. "Value as Any Object of Any Interest." In *Readings In Ethical Theory*. Second Edition. 138-151. Edited by Wilfred Sellars and John Hospers. New Jersey: Prentice-Hall.
- Pinker, Steven. 1997. *How the Mind Works*. London: Penguin.
- Plato. 1914. *Euthyphro, Apology, Crito, Phaedo, Phaedrus*. Translated by H.N. Fowler. London: William Heinemann.
- Prichard, H.A., 1912. "Does Moral Philosophy Rest on a Mistake." *Mind* 21:81. 21-37. ISSN: 0026-4423
- Putnam, Hillary. 1962. "It Ain't Necessarily So." *The Journal of Philosophy* 59:22. 658-671. DOI: 10.2307/2023596.
- Quine, W.V., 1951. "Main Trends in Philosophy: Two Dogmas of Empiricism." *The Philosophical Review* 60:1. 20-43. 10.2307/2181906.
- Rachels, Stuart. 2000. "Is Unpleasantness Intrinsic to Unpleasant Experiences?" *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 99:2. 187-210. DOI: 10.1023/A:1018622303499.
- Railton, Peter. 1986. "Moral Realism." In *Facts, Values and Norms: Essays Toward a Morality of Consequence*. 3-42. Cambridge: Cambridge University Press.
- Railton, Peter. 2017a. "Railton's Commentary." In *On What Matters: Volume 3*. 113-127. Authored by Derek Parfit. Oxford: Oxford University Press.
- Railton, Peter. 2017b. "Two Sides of the Meta-ethical Mountain." In *Does Anything Really Matter? Essays on Parfit on Objectivity*. 35-60. Edited by Peter Singer. Oxford: Oxford University Press.
- Rawls, John. *A Theory of Justice*. Oxford: Oxford University Press. 1973.
- Richards, Janet Radcliffe. 2000. *Human Nature After Darwin*. London: Routledge
- Ross, W.D., 1930/2002. *The Right and The Good*. Edited by Phillip Stratton-Lake. Oxford: Oxford University Press.
- Ruse, Michael. 1995. "Evolutionary Ethics. A Phoenix Arisen." In *Issues in Evolutionary Ethics*. 225-248. Edited by Paul Thompson. New York: State University of New York Press.
- Russell, Bertrand. 1964. *Human Knowledge: Its Scope and Limits*. New York: Simon and Schuster.
- Scanlon, Thomas. 2014. *Being Realistic About Reasons*. Oxford: Oxford University Press.
- Schroeder, Mark. 2008. *Being For: Evaluating the Semantic Programme of Expressivism*. Oxford: Clarendon Press.

- Sen, Amartya. 1979. "Equality of What?" In *The Tanner Lectures on Human Values*. Edited by Sterling McMurrin. 197-220. Cambridge: Cambridge University Press.
- Shafer-Landau, Russ. 2003. *Moral Realism: A Defence*. Oxford: Oxford University Press.
- Shafer-Landau, Russ. 2006. "Ethics as Philosophy: A Defence of Ethical Non-naturalism." In *Metaethics After Moore*. Edited by Terry Horgan and Mark Timmons. 209-232. Oxford: Oxford University Press.
- Shafer-Landau, Russ. 2012. "Evolutionary Debunking, Moral Realism and Moral Knowledge." *Journal of Ethics and Social Philosophy* 7:1. 1-37.
- Shafer-Landau, Russ. 2018. *The Fundamentals of Ethics*. Oxford: Oxford University Press.
- Sidgwick, Henry. 1907. *The Methods of Ethics*. Seventh Edition. London: Macmillan and Co.
- Singer, Peter. 1972. "Famine, Affluence and Morality." *Philosophy and Public Affairs* 1:3. 229-243. ISSN: 0048-3915
- Singer, Peter. 1993. *Practical Ethics*. Second Edition. Cambridge: Cambridge University Press.
- Singer, Peter. 2005. "Ethics and Intuitions." *The Journal of Ethics* 9. 331-352. DOI: 10.1007/s10892-005-3508-y.
- Singer, Peter. 2009. *The Life You Can Save*. New York: Random House.
- Singer, Peter. 2011. *The Expanding Circle: Ethics, Evolution, and Moral Progress*. Princeton: Princeton University Press.
- Singer, Peter and Katarzyna de Lazari-Radek. 2014. *The Point of View of the Universe: Sidgwick and Contemporary Ethics*. Oxford: Oxford University Press.
- Singer, Peter. 1975/2015. *Animal Liberation*. London: Bodley Head.
- Sinnott-Armstrong, Walter. 2011. "An Empirical Challenge to Moral Intuitionism." In *The New Intuitionism*. Edited by Jill Graper Hernandez. 11-28. London: Continuum.
- Slote, Michael. 1984. "Satisficing Consequentialism." *Proceedings of the Aristotelian Society* 58. 139-163. DOI: 10.1093/aristoteliansupp/58.1.139.
- Smith, Michael. 1994. *The Moral Problem*. Oxford: Blackwell.
- Sober, Elliott and David Sloan Wilson. 1998. *Unto Others: The Evolution and Psychology of Unselfish Behaviour*. Harvard: Harvard University Press.
- Spencer, Herbert. 1851. *Social Statics: Or the Conditions Essential to Human Happiness Specified and the First of them Developed*. London: John Chapman.
- Spencer, Herbert. 1884. *The Principles of Biology Vol.1*. London: Williams and Norgate.
- Spencer, Herbert. 1907. *The Data of Ethics*. London: Williams and Norgate.
- Stearns, Stephen C, and Rolf F. Hoekstra. 2005. *Evolution: An Introduction*. Second Edition. Oxford: Oxford University Press.

- Street, Sharon. 2006. "A Darwinian Dilemma for Realist Theories of Value." *Philosophical Studies* 127. 109-166. DOI: 10.1007/s11098-1726-6.
- Street, Sharon. 2008. "Reply to Copp: Naturalism, Normativity, and the Varieties of Realism Worth Worrying About." *Philosophical Issues* 18. 207-228. ISSN: 15336077.
- Street, Sharon. 2009a. "In Defence of Future Tuesday Indifference: Ideally Coherent Eccentrics and the Contingency of What Matters." *Philosophical Issues* 19. 273-298. ISSN: 0029-4624.
- Street, Sharon. 2009b. "Evolution and the Normativity of Epistemic Reasons." *Canadian Journal of Philosophy* 39. DOI: 10.1080/00455091.2009.10717649.
- Street, Sharon. 2010. "What is Constructivism in Ethics and Metaethics?" *Philosophy Compass* 5:5. 365-384. DOI: 10.1111/j.1747-9991.2009.00280.x.
- Street, Sharon. 2016. "Objectivity and Truth: You'd Better Rethink It." In *Oxford Studies in Metaethics*. 293-334. Edited by Russ Shafer-Landau. Oxford: Oxford University Press.
- Suikkanen, Jussi. 2019. "Contextualism, Moral Disagreement, and Proposition Clouds." In *Oxford Studies in Metaethics*. 47-69. Edited by Russ Shafer-Landau. Oxford: Oxford University Press.
- Temkin, Larry. 2017. "Some Reflections on Part Six of *On What Matters*." In *Does Anything Really Matter? Essays on Parfit on Objectivity*. 1-34. Edited by Peter Singer. Oxford: Oxford University Press.
- Tooby, John and Leda Cosmides. 1992. "The Psychological Foundations of Culture." In *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. 19-136. Edited by John Tooby, Leda Cosmides and Jerome H. Barkow. Oxford: Oxford University Press.
- van Vugt, Mark., Vladas Griskevicius and Wesley Schultz. 2014. "Naturally Green: Harnessing Stone Age Psychological Biases to Foster Environmental Behaviour." *Social Issues and Policy Review* 8, 1. 1-32. DOI:10.1111/sipr.12000
- de Waal, Frans. 2000a. *Chimpanzee Politics: Power and Sex Among Apes*. New York: Johns Hopkins University Press.
- de Waal, Frans and Flack, Jessica. 2000b. "Any Animal Whatever: Darwinian Building Blocks of Morality in Monkeys and Apes." In *Evolutionary Origins of Morality: Cross Disciplinary Perspectives*. 1-29. Edited by Leonard D. Katz. Thorverton: Imprint Academic.
- de Waal, Frans and Sarah Brosnan. 2003. "Monkeys Reject Unequal Pay." *Nature* 425. 297-299. DOI: 10.1038/nature01963.
- de Waal, Frans. 2006. "Morally Evolved: Primate Social Instincts, Human Morality, and the Rise and Fall of Veneer Theory." In *Primates and Philosophers: How Morality Evolved*. 1-58. Edited by Stephen Macedo and Josiah Ober. Princeton: Princeton University Press.

- de Waal, Frans, Zanna Clay, Klaus Zuberbühler and Lucie Ravaux. 2016. "Bonobos (*Pan paniscus*) Vocally Protest Against Violations of Social Expectations." *Journal of Comparative Psychology* 130:1. 44-54. DOI: 10.1037/a0040088.
- White, Roger. 2010. "You Just Believe That Because..." *Philosophical Perspectives* 24. 573-615. ISSN: 1520-8583.
- Wilberforce, Samuel. 1860. "[review of] *On the Origin of Species, by Means of Natural Selection; Or the Preservation of Favoured Races in the Struggle for Life.*" *Quarterly Review* 108. 225-264.
- Williams, Bernard. 1981. "Moral Luck." In *Moral Luck: Philosophical Papers 1973-1980*. 20-39. Cambridge: Cambridge University Press.
- Williams Bernard. 1981. "Internal and External Reasons." In *Moral Luck: Philosophical Papers 1973-1980*. 101-113. Cambridge: Cambridge University Press.
- Wilson. Edward O. 1995. "The Morality of the Gene." In *Issues in Evolutionary Ethics*. 153-164. Edited by Paul Thompson. New York: State University of New York Press.
- Wordsworth, William. 1850. "The Prelude or, The Growth of a Poet's Mind; An Autobiographical Poem." In *William Wordsworth: Complete Works*. East Sussex: Delphi.