Stockholm
University

Debunking Debunked?

Challenges, Prospects, and the Threat of Self-Defeat

Conrad Bakka

Doctoral Thesis in Philosophy at Stockholm University, Sweden 2023

Debunking Debunked?

Challenges, Prospects, and the Threat of Self-Defeat

Conrad Bakka

Academic dissertation for the Degree of Doctor of Philosophy in Philosophy at Stockholm University to be publicly defended on Friday 21 April 2023 at 13.00 in hörsal 11, hus F, Universitetsvägen 10 F.

Abstract

Metaethical debunking arguments often conclude that no moral belief is epistemically justified. Early versions of such arguments largely relied on metaphors and analogies and left the epistemology of debunking underspecified. Debunkers have since come to take on substantial and broad-ranging epistemological commitments. The plausibility of metaethical debunking has thereby become entangled in thorny epistemological issues. In this thesis, I provide a critical yet sympathetic evaluation of the prospects and challenges facing such arguments in light of this development. In doing so, I address the following central question: how could genealogical information undermine the epistemic justification of moral beliefs?

In Part I, I begin answering the central question by extracting explicit and implicit epistemic principles from three popular debunking arguments. These arguments, due to Gilbert Harman, Richard Joyce, and Sharon Street, generate principles concerning ontological parsimony, explanatory dispensability, epistemic insensitivity, lack of epistemic safety, unexplained reliability, epistemic coincidences, and explanatory constraints on rational belief. Having set out the principles tasked with explaining how genealogical information undermines, Part II of the thesis seeks to evaluate whether debunking arguments built on them succeed. To this end, I consider two types of challenges faced by such arguments.

First, there are strategies that attempt to block global moral debunking arguments. I argue that one popular such strategy, the so-called ‘third-factor strategy’, has been misunderstood. When understood correctly, it is of no help in answering debunking arguments. I then flesh out an alternative and more promising strategy for blocking such arguments. I then turn to internal challenges facing debunkers, particularly those who rely on ‘explanationist’ principles. I argue that explanationist debunking arguments, as well as most others, fall prey to one or more of four internal challenges: the implausibility of first-order epistemic principles, the threat of overgeneralization, the threat of self-defeat, and the need for costly metaepistemic commitments.

I conclude that current debunking arguments fail to establish that no moral belief is justified. By analyzing why existing arguments fail, I develop two conditions of adequacy that debunkers must satisfy in order to navigate the internal challenges successfully. I end by suggesting future directions that debunkers should pursue to rehabilitate the prospects for global moral debunking arguments.

Keywords: *evolutionary debunking arguments, genealogical debunking, reliability challenge, third-factor explanation, non-naturalism, explanationism, sensitivity, safety, conditional debunking arguments, metanormative, metaepistemology.*

Stockholm 2023
http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-215273

ISBN 978-91-8014-230-4
ISBN 978-91-8014-231-1

Department of Philosophy

Stockholm University, 106 91 Stockholm

DEBUNKING DEBUNKED?

Conrad Bakka

Stockholm
University

Debunking Debunked?

Challenges, Prospects, and the Threat of Self-Defeat

Conrad Bakka

ISBN print 978-91-8014-230-4
ISBN PDF 978-91-8014-231-1

Cover image: yx b, pexels.com

Printed in Sweden by Universitetsservice US-AB, Stockholm 2023

For Patrick

Acknowledgments

A thesis is a delicate flower. It starts its life as an unremarkable seed and only grows to realize its full potential given a nurturing environment. This thesis is no exception—it has been developed in a welcoming environment with the help of excellent co-gardeners who have contributed to making it bloom.

From idea until completion, the project has been subject to the guidance, encouragement, and excellent critical sensibilities of my supervisors, Jonas Olson and Björn Eriksson. They have been instrumental in weeding out all manner of imperfections. Conversely, what remains has benefitted immensely from their unrelentingly insightful comments. I am exceedingly grateful for their hard work.

During 2018–2019 I spent three and a half memorable months at the University of Groningen on a research visit. While there, I benefited greatly from the supervision of Bart Streumer and Daan Evers. Thanks also to the members of the metaethics reading group as well as the rest of the department for making my stay so enjoyable, both academically and otherwise.

Several people have provided valuable comments on one or more chapters of the manuscript. Thanks to Ainar Myiata, Andrea Asker Svedberg, Andreas Mogensen, Evelina Edfors, Jonathan Egeland, Malgorzata Michalowska, Mariona Eiren Miyata-Sturm, Nils Sylvan, Romy Eskens, Simon Allzén, Simon Knutsson, and Stina Björkholm. I want to especially thank Olle Risberg, Gunnar Björnsson, and Anandi Hattiangadi for providing valuable comments on the entire manuscript.

Much of the material from the thesis has been presented at various seminars and conferences, including the Stockholm University PhD-seminar in practical philosophy, *the Second Norwegian Workshop in Metaethics*, the University of Groningen graduate conference in philosophy, *Contemporary Issues Across Ethics & Epistemology*, the Tartu summer school on *Ethics, Empathy, and Errors,* as well as at a book symposium on Bart Streumer's *Unbelievable Errors*. At all these venues, I have received valuable feedback and I wish to thank all who attended.

The PhDs at the department, both past and present, visiting and permanent, have contributed immensely to creating a stimulating and cheerful work environment. Thank you all for making my years as a PhD student so enjoyable: Alice Damirjian, Amanda Thorell, Anders Lundstedt, Andrea Asker Svedberg, Daniel Ramöller, Eric Johannesson, Fabian Lorentzon, Felix Maier, Hana Kalpak, Henning Strandin, Henrik Ahlenius, Jimmy Goodrich, Johannes Nyström, Jonathan Egeland, Jonas Haeg, Karl Nygren, Lisa Hecht, Madeleine Hyde, Mats Ingelström, Meradjuddin Khan Oidermaa, Nils Sylvan, Olle Torpman, Romy Eskens, Sara Packalén, Sara Papić, Simon Allzén, Simon Knutsson, Stina Björkholm, and Tara Nanavazadeh.

I also wish to thank all other members of faculty and administration at the department for all the help, encouragement, and conversation throughout the years: Björn Brunnander, Emma Runestig, Erik Angner, Greg Bognar, Håkan Salwén, Isaac Taylor, James Nguyen, Krister Bykvist, Mattias Högström, Miira Tuominen, Mikael Janvid, Niklas Möller, Niklas Olsson Yaouzis, Orri Stefánsson, Peter Pagin, Robert Hartman, Sama Agahi, Siska De Baerdemaeker, Sören Häggqvist, Susanne Forssell, William Bülow O'Nils, Åsa Burman, Åsa Carlsson, and Åsa Wikforss.

A special thank you is due to my previous office mate, Jonathan Egeland, who, after putting up with all my questions about his own field of expertise, started working on the epistemology of debunking himself. That led to an insightful collaboration that made a significant positive contribution to the thesis, and in particular to the material in Chapter 7.

I would also like to extend a special thank you to my digital office mate, Kritika Maheshwari. From the depths of the pandemic and onwards, she provided a much-needed digital co-worker through all the ups and downs of the writing process. The production of this thesis would have been much less enjoyable in her absence.

Together with colleagues and friends, the generous funding I have received has enabled me to make the most of my time as a PhD student. Thanks to Erasmus+, Stockholm donation scholarships, Helge Ax:son Johnsons stiftelse, and the Department of Philosophy for their generous support.

I also wish to acknowledge the enormous debt I owe to my family for their care and support throughout my years as a PhD student. You have all been patiently waiting for me to explain the topic of my dissertation; maybe now it will finally become clear! Last, but in no way least, I am deeply grateful to Evelina for always being a reliable source of support and encouragement (and for the Sunday morning pancakes).

Contents

1 Debunking Arguments: An Introduction

1.1 Introduction

Recent decades have seen an explosion of interest in an old philosophical chestnut: When do the origins of a belief render it epistemically suspect? More specifically, under what circumstances should we give up a belief because we have become aware that it has an unfavorable origin? For a considerable amount of time, thinking that the origin (or genesis) of a belief could impact its epistemic credentials was considered a fallacy—the *genetic fallacy*. However, the “genetic fallacy” is not, in fact, a fallacy.[1]

More recently, the issue has been feverishly explored through the lens of evolutionary explanations of belief, and, in particular, of our moral convictions. It is no coincidence, many have thought, that such things as harm reduction, children, and social cooperation are of prime importance in our moral lives. It is exactly what they would be, if we were creatures belonging to a lineage whose ancestors’ moral attitudes had been significantly influenced by evolutionary selection pressures. Such pressures, the current line of thinking goes, inclined our ancestors toward having favorable attitudes regarding things that are closely linked to an increase in reproductive success, such as survival, the well-being of one’s children, and social cooperation.

Could such an evolutionary origin story for our moral beliefs, when fully spelled out and given empirical backing, provide us with information that would force us—on pain of irrationality or some other epistemic vice—to give them up? And if so, why should we think that evolutionary influences are the only form of murky origins that can undermine our beliefs? There are surely countless other subterranean and veiled influences that similarly affect what we believe—cultural and social affiliation, historical period, gender, religion, upbringing, and so forth. Does any influence from such factors similarly have the power to undermine our beliefs? Is showing that a belief has some of these influences among its causal or historical roots enough to force us to give it up?

[1] See Crouch (1993) for a historical overview of the purported fallacy. For arguments against its fallaciousness, see Sober (1994, 104–7), Joyce (2001, 159–61), and Klement (2002).

In broad strokes, such questions are what this thesis will set out to answer (although, in good academic fashion, I will relentlessly narrow its focus).

The following chapters will discuss and criticize several arguments—so-called moral debunking arguments—that attempt to leverage the origins of our moral beliefs in order to debunk them. I began this project with the ambition of charting the challenges facing such debunking arguments with the ultimate goal of formulating an argument that could successfully navigate them. As the work progressed, I found that moral debunking arguments face greater challenges than I previously thought. The project therefore evolved into an exploration of whether any such debunking arguments are likely to succeed at all. To that end, I have shown what would be required in order to construct a successful moral debunking argument, and the surprising theoretical burdens such an argument incurs.

For the impatient, let me already now reveal the ending. Through discussing a number of debunking arguments, I argue that any moral (or other) debunking argument needs to invoke substantial, first-order epistemological principles in order to explain how our beliefs are uniformly undermined. Such candidate principles are highly contested and often face challenges on purely epistemological grounds. The need for such commitments therefore drags a debunker out of the moral domain and into deep epistemological waters.

Even if a debunker were to secure a plausible first-order epistemological principle to power their argument, their work is far from done. I argue that debunking arguments that employ first-order epistemological principles tend to face a serious and underappreciated challenge: such arguments are either self-defeating or must go on to defend appropriate second-order epistemological commitments. To avoid self-defeat, proponents of debunking arguments will therefore need to take on highly contested and controversial commitments within a number of domains outside of the one they target. Depending on the type of argument in question, they can be required to commit to particular views about the epistemology of, not only the domain that they aim to debunk, but also of domains such as mathematics, modality, logic, as well as of metaphysics and epistemology itself.

Another surprising upshot of the following chapters is that there is a significant distinction between debunking arguments that seek to undermine our beliefs about some domain *full stop*, and those arguments that merely want us to change our minds about how to conceptualize a given domain. I will argue that the latter is by far the more promising type of debunking argument. However, *by its very nature*, this form of argument seeks to grant us true, justified beliefs about the domain(s) in question.

In short, I argue that the explanatory burdens facing proponents of debunking arguments have been seriously underappreciated. When properly understood, the theoretical package that is required risks rendering debunking arguments far less appealing and plausible than previously assumed.

In this introductory chapter, I begin by introducing the topic of the thesis—debunking arguments—and provide a brief overview of the literature on the topic (§1.2). I then provide a taxonomy of different types of debunking arguments as well as the necessary theoretical background and epistemological frameworks that I will rely on in the chapters that follow (§1.3). Subsequently, I discuss which accounts of our moral thought and talk have the most reason to worry about being subjected to debunking arguments (§1.4). In doing so, I also motivate certain restrictions to the scope of my central argument. Lastly, I provide an overview of the upcoming chapters and highlight my contributions to the frenziedly flourishing literature on debunking arguments (§1.5).

1.2 The Rise and Fall of Debunking Arguments

Imagine the following. Early one morning, you find yourself convinced that someone broke into your garden shed and stole your favorite flowerpot. After having your morning coffee, you recall that you had a particularly vivid dream that night about a thief stealing your beloved pot from your shed. Upon realizing the origin of your belief, you dismiss it with a chuckle.

Here, we have the materials needed to construct a simple debunking argument: information about the causal or historical origins of a belief—about its *genealogy*—that neutralizes that belief's epistemic credentials. You had a belief—that someone stole your flowerpot—and you had no reason to distrust it. However, you then realize that the belief originated from a notoriously unreliable belief-forming process: dreaming. Learning the source of your belief would seem sufficient to neutralize its epistemic credibility—sufficient, that is, to make it irrational, unjustified, or otherwise epistemically infelicitous for you to continue to hold on to it. Note that this is the case even if, unbeknownst to you and for entirely independent reasons, it nevertheless *were true* that someone stole your flowerpot.

Such simple scenarios show that it is hard to deny that a belief's genealogy can "debunk" it without showing it to be false. This might all seem painfully obvious to you—of course knowledge of a belief's origins can debunk it! So it has seemed to many philosophers as well. Arguments that appeal to the genealogy of our beliefs in order to debunk them have become increasingly prominent in many philosophical sub-disciplines. Such debunking arguments take many forms and differ, among other things, in their scope, target domain, and methodology.

Concerning the dimension of methodology, it will be useful to make a rough distinction between two types of debunking arguments and set one aside. The first type of debunking argument tends to be motivated by experimental results that purport to provide evidence of error in our philosophical judgments. Such arguments rely on experimental evidence showing the influence of irrelevant factors on our judgments as well as evidence of error and faulty reasoning. Based on such results, this type of argument concludes that our judgments about certain topics are epistemically defective. Call such arguments *experimental debunking arguments*.

Experimental debunking arguments have proliferated in the burgeoning fields of experimental philosophy and moral psychology.[2] This is in large part because experimental studies have indicated that people's judgments about hypothetical scenarios are sensitive to factors that seem entirely irrelevant to the questions at hand.[3] This comes out, for instance, when people are asked to make various philosophical judgments—often about ethics, metaphysics, and epistemology—when presented with hypothetical scenarios. People's answers to various questions about such hypothetical scenarios have been shown to be sensitive to factors such as cultural affiliation, the order in which cases are presented, differences in (seemingly philosophically irrelevant) wording and framing, choice of font, and even features of the physical environment, such as odor.[4] Some of these effects, such as the order of presentation, can persist even when the people asked are professional philosophers.[5]

Becoming aware that our philosophical judgments are significantly correlated with seemingly irrelevant factors such as these, some argue, should make us accept that the judgments are unreliable and consequently disallow them as a foundation for theorizing. For instance, Joshua Greene has argued on the basis of experimental evidence that certain normative judgments concerning

[2] For overviews of these fields and much of the relevant literature, see Tiberius (2014) and Alfano (2016).

[3] Many of the entries in Sytsma and Buckwalter (2016) contain an overview of this literature. Sauer (2018) provides a wide-ranging discussion of findings from experimental and moral psychology in the context of debunking arguments.

[4] See, respectively, Weinberg et al. (2001), Swain et al. (2008), Weigel (2011), Weinberg et al. (2012), Cecchetto et al. (2017). It's worth noting that some worries have been voiced about the methodological stringency of certain early experimental studies which found the type of effects mentioned in the main text (Woolfolk 2013).

[5] Schwitzgebel and Cushman (2015).

the rightness of actions—such as when to sacrifice people on trolley tracks—can track facts that appear to be morally irrelevant.[6] Such seemingly irrelevant facts include the degree to which the relevant actions involve being up close and personal with other people rather than, say, merely pressing a button.

Learning how these moral beliefs are formed—on the basis of irrelevant influences—can supply ammunition for an argument against retaining them. Since accounts such as Greene's hold that only particular ways of generating moral beliefs go wrong, a debunking argument built on it will target only certain pockets of philosophical or ethical theorizing.[7] While certainly interesting, I will not be discussing debunking arguments that target particular pockets of moral thinking based on experimental results. Even so, much of what I say will apply to them.

I will instead focus on the second class of debunking arguments. This type of argument grows out of a broader skeptical tradition and is concerned with larger epistemological worries that are, at least at first blush, far removed from experimentally testable results. Such arguments tend to start from abstract premises, often having to do with claims about causality, reduction, or explanation. Genealogical debunking arguments can, but need not, rely on empirical claims about the origins of our moral cognitive architecture or evolutionary influences on belief, but are rarely based on a concrete set of experimental findings of the type outlined above. Call the latter type of arguments *genealogical debunking arguments* (hereafter, simply debunking arguments).

This type of debunking argument is known under various names, such as *genealogical arguments*, *access problems*, *integration challenges*, *etiological arguments*, *isolation objections*, and *reliability challenges*.[8] To what extent the challenges going under these names are different, overlapping, or identical challenges is something of a research project in itself. However that may be, under this guise, debunking arguments have targeted beliefs about domains such as morality, mathematics, logic, modality, color, religion, time, causation, chance, reasons, and even ordinary mid-size perceptual objects.[9]

These two forms of debunking argument—experimental and genealogical—are primarily distinguished by the method(s) they employ and the reach of their conclusions. We will, through the coming chapters, investigate to what

[6] Greene et al. (2009); cf. Greene (2016).

[7] Similarly, Kelly (2011) argues that moral judgments that track or are otherwise based on feelings of disgust are unjustified.

[8] Korman (2019a) provides an excellent overview of such arguments.

[9] See, respectively, Joyce (2006), Field (1989, 25–30), Schechter (2010), Rea (2002, chap. 4), Schaffer (2019, sec. 2.2), Wilkins & Griffith (2012, sec. 6), Baron (2017), Price and Weslake (2008), Handfield (2016), Street (2006; 2009), Korman (2014).

extent empirical and experimental results are relevant to the formulation of successful genealogical debunking arguments. The relevance and importance of the distinction I have just drawn hang on the outcome of such questions. In any case, setting them out as distinct types of arguments will be helpful in setting out an overview of the literature on genealogical debunking arguments.

One of the most prominent targets for such debunking arguments is moral beliefs, and I will primarily be focusing on arguments targeting them. Even so, we will see that it is impossible to evaluate the success of debunking arguments targeting one domain without evaluating both their general features as well as whether they generalize to other domains.

One popular type of debunking argument aims to debunk beliefs by leveraging empirical facts about how evolutionary selection pressures have influenced and shaped our moral cognitive architecture, as well as our moral concepts, attitudes, and beliefs.[10] Call any such argument an *evolutionary debunking argument*.[11] This type of argument usually consists of two components. On the one hand, evolutionary debunking arguments tend to defend some empirical hypothesis, such as the claim that moral cognition is a biological adaption. On the other hand, they will argue that the empirical claim has implications for the epistemic credentials of moral beliefs.

I do not intend to argue for any empirical biological (or other) claims in what follows. For my purposes, it suffices to note the following. To the best of my knowledge, the hypothesis that moral cognition is a biological adaptation with the function of fostering social cooperation is the primary hypothesis across a wide range of disciplines, such as the evolutionary branches of psychology, biology, and anthropology.[12] Even so, the empirical claims made by evolutionary debunking arguments tend to go beyond whatever the abovementioned interdisciplinary consensus could reasonably be taken to amount to.[13] Sometimes, evolutionary debunking arguments are therefore made conditional on speculative empirical claims, such that a debunking argument succeeds *only if* its empirical speculations pan out.

The empirical details of various debunking arguments can be highly interesting and, in some cases at least, relevant for their epistemological upshots.

[10] Ruse (1986, chap. 6); Joyce (2001, chap. 6; 2006; 2016b); Lillehammer (2003); Kitcher (2005); Street (2006; 2009; 2011); Braddock (2017); Lutz (2018; 2020). For a historical overview of the influence of Darwinian theory on ethics, see Allhoff (2003).

[11] For overviews of the literature on such arguments, see Vavova (2015), Wielenberg (2016), and Korman (2019a).

[12] See, respectively, e.g. Cosmides et al. (2018), de Waal (2006), Tomasello (2016). For a dissenting view, see Arvan (2021).

[13] Cf. FitzPatrick (2020, sec. 2).

Despite this, a point that will be made repeatedly in what follows is that the importance and evaluation of evolutionary debunking arguments can often be judged in abstraction from most such empirical details. In fact, this is even admitted by many proponents of evolutionary debunking arguments.[14]

That being said, there has been significant philosophical debate, much of it speculative, over the exact nature of the purported evolutionary influence on both the cognitive architecture behind, and the content of, our moral attitudes.[15] In particular, it has seemed plausible to many that the overlap between the content of ordinary moral judgments and the evaluative attitudes that would enhance the evolutionary fitness of our ancestors (relative to individuals with different or no such attitudes) is not accidental.[16]

> [W]hy do we tend to judge that our survival is valuable, rather than worthless? Why do we tend to judge that we have special obligations to care for our children, rather than strangers or distant relatives? Why do we tend to view the killing of other human beings as a much more serious matter than the killing of plants or other animals?[17]

Many have argued that the best way to make sense of our moral and valuational practices is to accept that they have been significantly influenced by evolutionary forces. The general shape of an answer to the questions quoted above, on this type of account, has been forcefully set out by Sharon Street.

> Ancestors who made evaluative judgments of these kinds, and who as a result tended to respond to their circumstances in the ways demanded by these judgements, did better in terms of reproductive success than their counterparts.[18]

We will look in more detail at such empirical speculations concerning the relation between moral judgments and evolutionary fitness in Chapters 3 and 4. For now, it's sufficient to note that the type of debunking argument under discussion tends to rely on the claim that having certain evaluative tendencies was fitness-enhancing for our ancestors and that this, directly or indirectly, explains why certain patterns of moral judgments became widespread in the human population.

[14] Street (2006, 155); Joyce (2016a, 143; 2016b, 125). Klenk (2018, chap. 3) provides a sustained argument for this claim.

[15] Ruse (1986); Kitcher (2005; 2011); Joyce (2006); Street (2006).

[16] Hereafter, I will often drop the qualification that a trait, belief etc. confers a fitness advantage only *relative* to populations that lack that trait, belief etc. By 'fitness', I will hereafter mean such *differential reproductive success*.

[17] Street (2006, 132).

[18] Street (2006, 132).

Some might question the ultimate importance of this claim. Most of our cognitive architecture and dispositions, such as our capacity and tendency to form perceptual judgments about mid-size objects at medium range, have been fundamentally shaped by evolutionary forces.[19] Those who made reasonably veridical perceptual judgments avoided cliffs and fires and lived to see another day, while those who made significantly diverging perceptual judgments did not. In the case of perceptual judgments, we can expect that the fitness-enhancing effects conferred by making certain perceptual judgments rather than others are explained by the judgments being *true*.[20] Making systematically and grossly false perceptual judgments would not have conferred fitness-enhancing effects.

Proponents of evolutionary debunking arguments targeting moral beliefs argue that the selection pressures operating on our ancestors' moral attitudes were importantly different. It is not the case, the claim goes, that the evaluative tendencies in question were fitness-enhancing for our ancestors *because* they corresponded to moral reality. Instead, selection pressures favored those with pro-attitudes toward fitness-enhancing behaviors and practices *merely because* adopting such practices in itself conferred an increase in evolutionary fitness through fostering social cooperation and similarly beneficial behaviors.

If this is so, whether moral beliefs were true or false is immaterial to the question of whether evolutionary selection pressures would favor them. The patterns of moral judgment found in the human population would then be explained, not by an appeal to moral reality, but to the vagaries of whatever moral attitudes happen to have been fitness-enhancing. That moral beliefs were fitness-enhancing—to the extent that they were—is therefore to be explained by factors that are disconnected from the purported facts those beliefs are about. In short, the truth of the content of a moral belief would be irrelevant for the ability of the belief to confer a fitness advantage on a believer.

Compare again the case of perceptual beliefs. Such beliefs tend to be formed in response to causal stimuli. Perceptual beliefs therefore tend to *causally track* facts about our surroundings, in that the beliefs we form are responsive to the truths they are about. Our perceptual system has been selected for precisely because of its ability to track mid-size objects at a moderate distance. The truth or falsity of perceptual beliefs is *not* irrelevant for their ability to confer a fitness advantage. More generally,

[19] Stevens (2013).

[20] Cf. Mogensen (2014, 71–72). It might be possible, to some extent at least, to question such assumptions even in the perceptual case (Korman 2014; 2019b).

> [i]n order to explain why it proved advantageous to form judgements about the presence of fires, predators, and cliffs, one will need to posit in one's best explanation that there *were indeed* fires, predators, and cliffs.[21]

Moral beliefs, on the other hand, have been claimed to lack any such causal connection, or any other type of explanatory connection, to the purported moral facts they are about. If that is so, one might worry that moral beliefs do not track moral reality in any meaningful way. We believe that social cohesion is good, but it is hard to see how that belief is explained by there being some moral fact that shapes or spurs the belief. On the contrary, many have thought, we form the belief, not on the basis of grasping facts about the moral worth of social cohesion, but become disposed to form the belief merely because holding it is itself fitness-enhancing.

The worry is sometimes brought out by the claim that we would be equally likely to see the current pattern of moral beliefs in the human population, even if those beliefs were false.[22] This is sometimes presented as the claim that moral beliefs fail to 'track the truth', which is in turn used to support the conclusion that moral beliefs are unreliable. This rough line of argument encapsulates how moral beliefs have often been claimed to be undermined by their evolutionary genealogy.[23]

Exactly how these claims are to be understood, including what it means for moral beliefs to 'not track' moral facts, be unreliable, or undermined, and how one gets from one to the other, varies between authors, or is sometimes left entirely unexplained. The crux of such arguments, we will discover, is how they move from the availability of evolutionary explanations of moral belief to whatever epistemological conclusion they are championing.

The above paragraphs have covered a lot of ground and have glossed over deep and complicated issues in metaethics, epistemology, and evolutionary theory. In the chapters that follow we will return to fill in these details.

Evolutionary debunking arguments have been met with many responses, some of which will be discussed in more depth in later chapters. A set of such responses have questioned or denied the empirical claim that evolutionary processes have significantly influenced our moral beliefs.[24] Others have accepted such influence, as well as accepting that moral facts are irrelevant for explain-

[21] Street (2006, 160 fn. 35).

[22] Ruse (1986, 254); Joyce (2001, 163); Sinnott-Armstrong (2006, 43); Braddock (2017).

[23] Cf. Kahane (2011); Vavova (2015). See Egeland (2022) for an attempt at capturing this, and many other ways, of formulating debunking arguments in a single argument schema.

[24] Parfit (2011, sec. 119); Nagel (2012); FitzPatrick (2015), Huemer (2016); Isserow (2019).

ing the fitness-enhancing effects of moral beliefs, but denied that this necessarily leads to moral skepticism.[25] One version of this response claims that when we consult our intuitions and consider what the moral facts are, we can be confident that we have gotten it right, whether or not evolutionary processes have disposed us to have the intuitions we do.[26]

Another line of response has been to claim that moral beliefs *do* track moral truths. This has been suggested by some who take moral facts to be neither empirically accessible nor causally efficacious. Such theorists claim that believing in what we can call *non-natural moral facts* could be fitness-enhancing *because of* the truth of the content of such beliefs.[27] Alternatively, we might have general rational capacities that have led us to believe non-natural moral truths *because* they are true, though it is not the truth of such beliefs that makes them fitness-enhancing.[28] Others, who claim that moral facts are empirically accessible and/or causally efficacious—i.e., that they are *natural facts*—have argued that our moral beliefs could thereby track the relevant moral facts.[29]

One last line of reply, of particular interest for our purposes, is the so-called third-factor strategy. This strategy aims to grant most of what a debunker is claiming, including the evolutionary influence on our moral attitudes as well as the irrelevance of moral facts for explanations of the fitness effects of moral beliefs. The third-factor strategy nonetheless tries to secure the explanatory connection between moral facts and moral beliefs that debunkers claim is missing.[30] The strategy does this by appealing to a form of indirect explanatory connection overlooked by earlier debunking arguments.

The claim is that there could be a factor that participates in the explanation of *both* our moral beliefs as well as moral facts, although there is no direct explanatory connection between the latter two. This would fend off worries about moral beliefs being “off-track” or unreliable, as it would secure a reliable and explainable overlap between the content of our moral beliefs and the moral facts. In response to the third-factor strategy, debunkers have begun

[25] For discussion of such strategies, see White (2010, 589), Srinivasan (2015, 347–49), Mogensen (2015), Clarke-Doane (2015; 2016; 2020), and Korman (2019a, 6).

[26] Dworkin (1996, 125–27); Parfit (2011, 531). For discussion, see Vavova (2014).

[27] Parfit (2011, 534–38).

[28] FitzPatrick (2015); Huemer (2005, 99–102; 2016).

[29] Copp (2008); Lott (2008). This definition of what it is for a fact to be natural is quick and dirty but suffices for our purposes. For discussion, see Lutz and Lenman (2021, sec. 1.1).

[30] For accounts that are often classified as third-factor explanations, see Copp (2008), Enoch (2010; 2011, chap. 7), Wielenberg (2010; 2014, chap. 4), Schafer (2010), Brosnan (2011), Skarsaune (2011), Behrends (2013).

making their arguments more precise, by working out explicit and detailed epistemological principles intended to rule out such third-factor replies.[31]

This has led to the literature on evolutionary debunking arguments recently refocusing on two questions: What epistemological principle explains how evolutionary explanations debunk moral beliefs? And could such a principle rule out replies to debunking arguments, such as the third-factor strategy? The ensuing discussion has driven the literature on moral debunking arguments in directions that go far beyond their initial concerns. Let me now list four central issues of this kind that will serve as focal points for the coming chapters.

Epistemological principles. Debunking arguments, I will argue, should be understood as ultimately relying on first-order epistemological principles either implicitly or explicitly. As we will see repeatedly throughout the coming chapters, when spelled out, any such candidate principle is hotly contested on purely epistemological grounds. Additionally, the more replies to debunking arguments a debunker seeks to rule out with a particular epistemological principle, the more radical, and controversial, that principle will need to be. If a debunker opts to embrace an epistemological principle that is capable of ruling out replies like the third-factor strategy, they will thereby open themselves up to new avenues of purely epistemological criticism.

(Over)generalization. Most debunking arguments target a particular type—or at most a few types—of belief, such as moral or religious belief. The arguments are intended to show that such beliefs are epistemically deficient in a way that other beliefs are not. The candidates for being the operative epistemological principle in debunking arguments are often quite general. This incurs a risk of such debunking arguments becoming *too* successful. Such principles, when employed in a debunking context, can result in too many of our beliefs being debunked—including beliefs few of us are willing to give up.

In particular, debunking arguments risk becoming instances of global skeptical arguments, which few debunkers wish to defend. Even worse, we will see that when debunking arguments are given a global reach, they are likely to be self-defeating. On the flip side, if the principles are weakened to avoid such results, they risk losing the ability to make us give up even the beliefs initially targeted by the debunking argument.

Self-defeat. Certain ways of construing debunking arguments have been thought to lead debunkers to straightforward self-defeat.[32] Whatever epistemological principle a debunker employs must avoid this outcome. There is,

[31] Lutz (2018; 2020); Korman and Locke (2020; 2021); cf. Faraci (2019).

[32] Pust (2001); Vavova (2014); Srinivasan (2015); Kyriacou (2016).

for instance, a danger that the moral domain, concerning what we have (moral) reason to do, is sufficiently similar to the epistemic domain, concerning what we have (epistemic) reason to believe, that targeting one risks targeting the other. This could lead a debunker to target the epistemic status of the premises contained in their own argument. There is again a tension in finding a principle that is sufficiently strong to undermine moral beliefs, without also indiscriminately afflicting neighboring domains, such as epistemic beliefs.

Metaepistemology. Epistemology is concerned with first-order questions concerning, among other things, the nature of knowledge and justification. Metaepistemology, on the other hand, is concerned with second-order questions about, among other things, knowledge and justification. A central metaepistemological question is therefore: What is the nature of epistemic facts?[33] Metaepistemological issues have not received the scrutiny that their metaethical counterparts have. As a result, the metaepistemological landscape is far murkier. This is starting to be rectified and such questions are currently seeing a sprawling literature with myriad proposals for how to understand our epistemic thought and talk, as well the nature of epistemic facts themselves.[34]

Debunkers, I will argue, have explicitly or implicitly been committed to epistemological principles that, at least *prima facie*, undermine the epistemic status of premises of their own arguments. To avoid this outcome, as well as navigate the three challenges above, debunkers must move beyond a commitment to first-order epistemological principles. In addition, they will need to stake a claim in metaepistemological debates over the nature of epistemic facts. Only by doing so can they avoid their own arguments being vulnerable to self-defeat. This involves charting unexplored ground and therefore involves taking on a large explanatory burden.

These four issues, taken together, pose a significant obstacle to constructing a debunking argument that both employs a plausible epistemological principle and is capable of ruling out various replies to debunking arguments.

Early debunking arguments often left the epistemic details required to answer these challenges unsaid, and usually only hinted at them by the use of scenarios, metaphors, and analogies. As debunkers have made their arguments

[33] In general, I will use ‘epistemic’ to refer to attitudes, facts and states related to knowledge and justification and the like, while using ‘epistemology’ and ‘epistemological’ to refer to the outputs of theorizing about epistemic phenomena. That S knows that p is therefore an epistemic fact, while the claim that knowledge implies true, justified belief is an epistemological principle.

[34] For a sampling of this literature, see Cuneo (2007), Greco (2015), McHugh et al. (2018a), McHugh et al. (2018b), Cowie (2019), Carter and McKenna (2021).

more explicit, they have come to make controversial and broad-ranging epistemological commitments that make their arguments less appealing.

Because of such difficulties, the tide seems to have shifted somewhat against evolutionary debunking arguments. It has been increasingly common to see counter-arguments to the effect that non-skeptics—realists and anti-realists alike—have the resources to block debunking arguments. Some go even further and claim that there are structural features of evolutionary debunking arguments that make them unworkable, even in principle.[35]

The central question to ask about genealogical debunking arguments targeting moral beliefs is therefore two-fold:

(i) Which epistemological principle(s) could explain how genealogical information about moral beliefs could suffice to debunk them?

(ii) In light of the answer to (i), can debunking arguments avoid the threat of overgeneralization and self-defeat?

The four central issues set out above, together with these two questions, form something of a rough outline of the topics to be addressed in later chapters.

Let us now turn to the theoretical background and assumptions that will inform the discussion in later chapters.

1.3 Moral Debunking: Taxonomy and Epistemology

An important lesson from the past fifteen years of work on debunking arguments is that it has suffered from a lack of attention to its epistemological commitments and presuppositions. Employing the somewhat vague terms ‘debunking’, ‘undermining’, and ‘off-track’, as I have myself done so far, has led to an overestimation of how easy it is to secure a thoroughgoing moral skepticism on the back of genealogical explanations of moral beliefs. How are we to understand the epistemological underpinnings of debunking arguments? And what is it to “debunk morality”?

We can categorize the most common ways in which ‘debunking’ is used in the context of morality by borrowing, and slightly modifying, a taxonomy set

[35] Srinivasan (2015); Kyriacou (2016).

out by Richard Joyce.[36] Arguments attempting to "debunk morality" typically aim to establish one or more of the following conclusions:

1. *FALSE*: All moral judgments are false.[37]

2. *UNJUSTIFIED:* All moral judgments are unjustified.

    2a. *ALWAYS*: No moral judgment has ever been justified.

    2b. *NOW*: Some moral judgments are (or have been) defeasibly justified, but all moral judgments would be (or have been) undermined by information about their genealogy.

3. *NO KNOWLEDGE*: No one possesses moral knowledge.

4. *CONDITIONAL*: Certain metaethical theories (e.g., moral realism)—but not all such theories—should be rejected.

5. *LOCAL*: Certain moral theories (e.g., Kantianism)—but not all such theories—should be rejected.

I will briefly say something about each of these conclusions that debunkers might pursue. Few debunkers have explicitly argued for *FALSE*, claiming that the genealogy of our moral beliefs, on its own, allows us to conclude that such beliefs are all false.[38] Even so, some argue that genealogical information shows that our moral beliefs are *likely* to be false.[39]

Genealogical information alone rarely licenses the conclusion that a given belief is false. Rather, it more typically shows that whatever grounds we had for holding the relevant beliefs are invalidated. Recognizing this, debunkers often argue instead for some version of *UNJUSTIFIED*, holding that genealogical information shows that our moral beliefs uniformly fail to be justified, which leaves it open whether they are in fact true or false.[40]

[36] Joyce (2016a, 143).

[37] The target set of judgments (i.e., *all* moral judgments) might have to be stated a bit more carefully. Depending on how one defines 'moral' one might need to restrict it to all non-tautological moral beliefs. We will presumably remain justified in believing that 'either murder is wrong or murder is not wrong'. I will hereafter disregard any such qualifications.

[38] Some interpret Ruse (1986) as arguing for this (cf. Joyce 2016a, 144).

[39] Joyce claims that a debunking argument, together with traditional metaethical claims, can show moral beliefs to be "probably false" (2001, 168; cf. 2006, 244 n17). Similarly, Street claims that, if moral realism is true, moral beliefs are "likely to be false" (Street 2006, 125).

[40] Joyce (2006); Braddock (2017); Lutz (2018; 2020); Korman and Locke (2020).

While the importance of our moral beliefs being false is perhaps immediately clear, what is the importance of the justificatory status of our beliefs? For one thing, epistemic justification is intimately connected to a number of things we care deeply about: it is commonly taken to be a necessary condition on knowledge as well as on permission and obligation to believe.[41] If our belief that p is not justified, on such accounts, we neither know that p, ought to believe that p, nor are we permitted to believe p.[42] Even if justified belief were not required for knowledge, it would still be an important concept in its own right, in part for its importance for rational belief.[43] If no moral belief is justified, it would therefore seem that our assertoric moral practice would be under serious threat.[44]

Since debunking arguments often conclude that moral beliefs are uniformly unjustified, it becomes important to clarify what the relevant notion of ‘justification’ is. At the most general level, debunking arguments are concerned with targeting *epistemic justification*—roughly, reasons for holding a belief—as opposed to practical or pragmatic justification. A second distinction concerns to what we apply the property of justification, propositions or doxastic states. This is the distinction between propositional and doxastic justification.[45]

Propositional justification concerns whether you can justifiably believe a proposition, whether or not you actually believe it. I am justified in believing that the sun will rise tomorrow, in this propositional sense, because of the inductive evidence I possess in favor of it being true. I would be propositionally justified in believing this, even if I did not in fact hold the belief.

Doxastic justification, on the other hand, concerns whether a belief that is in fact held by an agent, is held justifiably. Doxastic and propositional justification can come apart. Say I believe that the sun will rise tomorrow, but only because I believe that the deity Helios rides across the sky and lights the sun each morning. In this case, I have propositional justification for the content of my belief, in virtue of the inductive evidence available to me. I do not, however, possess doxastic justification for my belief that the sun will rise tomorrow, since the belief is *based* on grounds that do not sufficiently support it.

[41] For former claim, see Ichikawa and Steup (2017, sec. 1.3). For the contrary view, see Kornblith (2008). For the latter claims, see Alston (1988).

[42] One might reasonably think justification comes in degrees. In what follows, I will mostly ignore this complication.

[43] Kvanvig (2003, chap. 3).

[44] The essays in Garner and Joyce (2019) map out some possible consequences of moral skepticism (in the guise of moral error theory) for moral practice.

[45] For critical discussion of the distinction, see Turri (2010).

In what follows, we will be concerned with each type of justification at different points. I will sometimes make this explicit, although I will allow myself to use formulations that sometimes blur the line between them when nothing hangs on the distinction.

Beyond the above distinctions, epistemic justification is a deeply contested notion, and there is little agreement surrounding it. In particular, much hangs on one's allegiance along the internalist-externalist axis of views about the nature of justification.[46] This presents debunking arguments arguing for *UNJUSTIFIED* with both a substantial and presentational hurdle, which I will return to below. Importantly, it can affect whether a debunking argument has *UNJUSTIFIED ALWAYS* or *UNJUSTIFIED NOW* as its conclusion. Let me illustrate.

Epistemic externalists hold that facts about justification can, wholly or partly, depend on facts that are not internally accessible to us. The most popular form of externalism, *reliabilism*, requires for (doxastic) justification that, very roughly, a belief has been formed by a reliable belief-forming mechanism.[47] If genealogical information would show that moral beliefs were produced by an *unreliable* belief-forming mechanism, such beliefs would not (now or ever before) satisfy a necessary condition for being justified. Hence, moral beliefs would never have been justified to begin with. It is therefore open to various types of externalists, such as reliabilists, to argue for *UNJUSTIFIED ALWAYS*.

Certain accounts of justification, such as simple forms of reliabilism, cannot account for defeasibly justified moral belief. This means that no belief can be such that an agent is justified, at a time, in believing it, but where that status is lost at a later time. This is often taken to count against such views. The idea that justification is usually, if not always, defeasible in this sense, is commonly recognized as an essential part of contemporary epistemological theory.[48] Externalists will therefore usually want to accommodate this possibility, or at least something very much like it.

Below I will outline a general framework of defeasible justification in the context of internalist views of epistemic justification. Such a framework, at a sufficient level of abstraction, may ultimately be acceptable to both internalists and many externalists, though the details may be spelled out differently.

[46] For an overview of the internalist-externalist debate, see BonJour (2010).

[47] The classical statement of reliabilism is found in Goldman (1979).

[48] Kvanvig (2007); Schechter (2013b, 437).

If externalists were to take on board some such account of defeasible justification, they could accept debunking arguments that argue for *UNJUSTIFIED NOW*.[49]

Epistemic internalists hold, roughly, that facts about justification supervene exclusively on states that are internally accessible to the agent.[50] Simplified, the internalist sense of justification concerns whether a belief is rationally held by an agent, as evaluated from the agent's perspective. I will now briefly outline two ways in which such a theorist might wish to understand the epistemological underpinnings of debunking arguments in terms of defeasible justification. I'll discuss these two frameworks—undercutting defeat and higher-order defeat—in turn.

As long as we are not global skeptics, we take some of our beliefs to be justified. That your belief that p is justified is not always a permanent status; justification is usually, if not always, *defeasible* in that a previously justified belief can lose that status.[51] The conditions that would defeat the justification for a belief are known as *epistemic defeaters*.[52] Given that such conditions obtain with respect to your belief that p, you have a defeater for the justification for your belief that p.[53] For simplicity, I'll say that you then have a defeater for the belief that p. Someone who is defeasibly justified in believing that p at one time might therefore become unjustified in maintaining that belief at a later time given that the agent comes to possess a defeater for the belief.

Roughly, an epistemic defeater can then be described as a reason to withhold or give up a belief where, had it not been for the defeater, one would have

[49] Externalist views that can accommodate defeasible justification include those that incorporate some traditionally internalist commitments, such as evidentialist reliabilism (Comesaña 2010; Goldman 2011).

[50] E.g. Conee and Feldman (2001). This might plausibly need to be restricted to the agents *non-factive* mental states, where a non-factive mental state is one where its propositional content is not necessarily true.

[51] On indefeasible justification, see Grundmann (2011, 162–63).

[52] My outline below owes much to Sudduth (n.d.) and Grundmann (2011). For a detailed framework of epistemic defeat, see Pollock (1986).

[53] Much controversy surrounds the exact nature of defeaters, which I will mostly attempt to steer clear of. For instance, a view I will not discuss here, but which could sensibly be added to the above taxonomy, takes true propositions outside of the agent's perspective to be defeaters, not for justification, but for *warrant* (Plantinga 2000, 359). Warrant, on such accounts, is understood as what turns a true belief into knowledge. A defeater for p, on such a view, would be a fact such that, had it not obtained, a true belief that p would have been knowledge. In other words, true, unknown propositions can block warrant. This type of "defeat" does not involve neutralizing prior defeasible justification, and one can therefore ask in what sense it's meaningfully called 'defeat' (Grundmann 2009).

been justified in holding the belief.[54] The reason in question is *prima facie* because an epistemic defeater might itself be defeated.[55] For instance, if currently available genealogical explanations of moral beliefs are a defeater for those beliefs, but it turns out that we have been misled in our assumptions about the genealogy of our moral beliefs, then we might in turn have a defeater for the former defeater (a so-called *defeater-defeater*). Strictly speaking, it is therefore only undefeated defeaters that neutralize the justification for a belief.

Defeaters come in different types, of which two of the standardly recognized types are undercutting and rebutting defeaters.[56] An *undercutting defeater* for a belief that p is a reason for giving up that belief, without being a reason to believe that p is false. Such undercutting defeaters are reasons for thinking that whatever grounds one had for believing that p are defective.[57] Consider again the case where you believe that someone has stolen your favorite flowerpot from your garden shed. Upon finding out that your belief originated in a dream, that new piece of information constitutes a reason to think that the grounds you had for holding the belief are really no indication of its truth at all. As such, it is a reason for giving up the belief, without being a reason for thinking that the belief is false. After all, it could still be the case that someone *has* stolen your favorite flowerpot, but any grounds you had for believing so have been invalidated.

A *rebutting defeater* for a belief that p is a reason to hold p to be false or to hold some incompatible proposition q to be true.[58] For instance, if your neighbor tells you that they saw someone take off with a flowerpot from your garden shed, you are plausibly justified in believing this to be so. However, if you go into your shed and find that your beloved pot is still there, this would be a reason to think that it is *not* the case that anyone stole your pot. That no pot appears stolen would therefore be a rebutting defeater for the defeasibly justified belief that someone stole your pot.

Let us now apply the framework of defeat to moral debunking arguments, which results in the following picture. Roughly, despite many of our moral beliefs initially being defeasibly justified, the new information we receive by considering their genealogy shows them to be unreliable (by being off-track, explanatorily irrelevant or whatever), or to suffer some other epistemic defect,

[54] Exactly how to formulate defeaters are a matter of controversy. See Moretti and Piazza (2018) for an overview of the issues involved.

[55] Grundmann (2011, 161–62). Hereafter, I will often drop the *prima facie* qualification.

[56] Some taxonomies of defeaters also add further ones (e.g. Grundmann 2011, 158).

[57] Pollock (1986, 39).

[58] Pollock (1986, 38).

which constitutes a defeater for those beliefs. Exactly what constitutes the defeater will depend on a number of issues. For instance, focusing on lack of reliability, an externalist might consider the lack of reliability *itself* to constitute a defeater, while an internalist might hold that it is *acknowledged* lack of reliability that does so.[59] In the latter case, after gaining the relevant information, we would no longer be justified in maintaining our moral beliefs. In this way, many debunkers have argued for *UNJUSTIFIED NOW*.[60]

Another way in which debunkers might argue for *UNJUSTIFIED NOW* is by holding that the genealogical information about our moral beliefs provides us with evidence about our evidence. For instance, consider phenomenal conservatism. This view of epistemic justification holds that if it seems to you that p, then, absent any defeaters, you are defeasibly justified in believing that p.[61] Our beliefs would therefore be justified in virtue of our (undefeated) intuitions. However, if we come to learn that these intuitions are the product of an unreliable process, we might then possess higher-order evidence—evidence about our evidence—that undermines the first-order evidence provided by those intuitions.

For instance, if it seems to me that God exists, but I come to learn that religious convictions arise out of wishful thinking, then I've acquired evidence about my evidence that indicates that the evidence was gained through an unreliable process.[62]

Defeat that occurs in this way—i.e., *higher order-defeat*—might seem similar to the account of epistemic defeat set out above, but some have argued that higher-order defeat is different in kind from both undercutting and rebutting defeat.[63] Depending on whether this is so, an account of debunking arguments in the framework of higher-order defeat might either be identical or different from the one set out in terms of undercutting defeat above.[64] Much

[59] Moretti and Piazza (2018, 2846).

[60] Mogensen (2014); Joyce (2016a, 157); Braddock (2017); Lutz (2018; 2020); Korman and Locke (2020). For a critical discussion of the framework(s) of epistemic defeat in the context of moral debunking, see Klenk (2018; 2019). For examples of the framework of defeat being applied in debunking arguments targeting other domains than morality, see Plantinga (1993, chap. 12), Merrick (2003), Thurow (2013), and Barker (2020).

[61] Huemer (2007, 30).

[62] Freud (1927), in an early debunking argument, famously or infamously, claimed that religious conviction arises out of wishful thinking.

[63] Feldman (2005); Christensen (2010); Schechter (2013b); Lasonen-Aarnio (2014); DiPaolo (2018). For the opposing view, that higher-order defeat defeats, when it does, through undercutting defeat, see Risberg and Tersman (ms.; cf. 2020).

[64] For discussion of the relation of higher-order defeat to debunking arguments, see the entries in Klenk (2020).

more could be said about these frameworks, and some more will be said later, but this should provide a sketch of the various ways in which genealogical information could render moral beliefs unjustified on a wide variety of epistemological frameworks.

Return now to the taxonomy of epistemological conclusions pursued by debunkers. *FALSE*, as well as some versions of *UNJUSTIFIED*, entail *NO KNOWLEDGE*. At least, this is so on the assumption that justification is required for knowledge. However, as Gettier showed, one can have true, justified belief that does not amount to knowledge.[65] One could therefore argue for *NO KNOWLEDGE* directly, without arguing for either *FALSE* or *UNJUSTIFIED*. Relatively few debunking arguments have been formulated this way, although discussion of debunking arguments, or replies to them, sometimes takes this form.[66] One way in which this type of debunking argument could be formulated is by arguing that our moral beliefs fail to satisfy a necessary condition on knowledge. We will consider such arguments in Chapter 6.

Some debunkers do not seek to ultimately establish a skeptical conclusion along the lines of *FALSE*, *UNJUSTIFIED*, and *NO KNOWLEDGE*. Instead, they argue that such skeptical conclusions hold true only relative to certain conceptions of our moral thought and talk, and in particular, various forms of moral realism. *Moral realism*, as I will use the term, denotes any view that takes there to be stance-independent moral truths, where these are truths that do not obtain in virtue of our attitudes or beliefs about them.[67] Call this type of debunking argument, which argues for some version of *CONDITIONAL*, while denying *FALSE*, *UNJUSTIFIED*, and *NO KNOWLEDGE*, *conditional debunking arguments*.[68]

Such conditional arguments have a two-part structure, consisting of one conditional claim and one claim about theory choice. The conditional claim has the following structure: If a certain metaethical view (e.g., moral realism) is true, then evolutionary explanations of moral belief result in a skeptical upshot (e.g., that our moral beliefs are all unjustified).

Upon defending the conditional claim when applied to e.g., moral realism, one is faced with a choice. It is still possible to endorse moral realism, but one would then be forced to accept that all moral beliefs are unjustified (i.e., accept some version of *UNJUSTIFIED*). Such *skeptical realism*, a proponent of a con-

[65] Gettier (1963).
[66] Wielenberg (2010); Brosnan (2011); Tropman (2014).
[67] Cf. Shafer-Landau (2003, 15).
[68] Cf. Korman (2019a, 3–4).

ditional debunking argument will argue, is unpalatable. Instead, the conditional debunker will claim, one should take the conditional debunking argument to tip the scales in favor of some alternative, non-skeptical, anti-realist metaethical theory that is not threatened by the debunking argument. We will look at one conditional debunking argument of this form in Chapter 4.

Debunking arguments vary in their scope. Some target all beliefs within their target domain. Evolutionary debunking arguments targeting moral belief are often *global* in this sense, such as *FALSE*, *UNJUSTIFIED*, *NO KNOWLEDGE*, and *CONDITIONAL*.[69] Some debunking arguments have a more selective, *local* target and only seek to undermine a subset of beliefs within a domain. This type of argument is captured by *LOCAL* in the taxonomy. An example is de Lazari-Radek and Singer's evolutionary debunking argument against deontological moral beliefs, which is intended to leave consequentialist moral beliefs unscathed.[70]

Although I will not argue for it here, I believe such local debunking arguments are very hard to keep contained.[71] As we will see, even global debunking arguments are difficult to contain within one domain, as they tend to generalize beyond the domain they are initially intended to target.[72] I will therefore largely restrict my focus to global debunking arguments targeting moral beliefs. Even so, I will at various points discuss how, and if, such global moral arguments generalize in a way that forces them to target non-moral beliefs as well, such as mathematical or non-moral philosophical beliefs.

We have now considered a number of conclusions defended by debunkers as well as how they interact with various views of epistemic justification. As far as possible, I will try to gloss over issues that arise from adhering to some particular account of epistemic justification, and I will use the term 'undermine' as a placeholder for whatever preferred account one has of how a belief is rendered unjustified, whether it is through undercutting defeat, higher-order defeat, or by the belief never having been justified in the first place, or by some other account.

This is not ideal, but the alternative would involve dedicating a substantial and unrelenting focus on issues that are at least somewhat tangential to what I

[69] Joyce (2006); Street (2006).

[70] de Lazari-Radek and Singer (2012). A similar argument is championed by Greene (2008; 2013).

[71] Cf. Kahane (2014); Rini (2016). We will see one example of how a local debunking argument could be motivated in §6.3.2.

[72] Srinivasan (2015); Vavova (2014); Kyriacou (2016).

take to be the central issues at hand. In addition, it would likely involve choosing one framework, and losing the ability to say something general about issues that should be of relevance across such epistemological frameworks.

Two drawbacks to this decision should be mentioned. First, one will have to translate the relevant epistemological principles and claims I discuss into one's preferred epistemological framework. Sometimes I will suggest how this can be done, but often I will not. Second, it might be that, when translated into some such framework, further issues will arise about how, exactly, the mechanics of the relevant framework allows for debunking to occur.[73] Such issues, while highly interesting, are not ones I will engage with here to any serious degree. Even so, I believe the issues I will discuss should, to a great extent, be relevant across different frameworks.

An important element of any evolutionary debunking argument is to whom the argument is directed and ultimately constitutes a threat to. I turn to this question in the next section.

1.4 Who's Afraid of Being Debunked?

The debunking arguments I am primarily concerned with target moral beliefs. Different accounts of moral thought and talk might have different tools available for rebutting such arguments. It used to be something of a received view that debunking arguments pose the steepest challenge for robust moral non-naturalist realists (hereafter *moral non-naturalists*).[74] Conversely, it's been commonly thought that moral naturalists have a straightforward strategy available for avoiding debunking arguments, as they can appeal to various dependency relations (identity, reduction, constitution, grounding) in order to explain how our moral beliefs are in some way closely related or connected to natural facts.[75]

The consensus on these issues seems to be rapidly dwindling. There is now something of a free-for-all as to claims about which metaethical conception faces the gravest threat from debunking arguments. Some claim that only (and all) naturalists need worry about them.[76] Others, that moral quasi-realists are

[73] For a sustained investigation of such issues with respect to the framework of defeat, see Klenk (2018).

[74] Enoch (2011, 160); Joyce (2006, 209–10); cf. Schechter (2018a, 443). Enoch (2011) coined the name "robust realism" and defends moral non-naturalism as part of a broader non-natural normative realism.

[75] Enoch (2011, 160); Joyce (2016b, 126–27).

[76] Bogardus (2016).

no less targeted than moral realists.[77] Still others, that all non-skeptical metaethical views that accept that moral judgments can be correct or incorrect face a serious challenge from it.[78] It has even been suggested that non-cognitivism generally might face some form of debunking challenge.[79]

Discussing and drawing the differential implications of various debunking arguments for every metaethical view under the sun would be daunting, and I will therefore need to restrict my scope. For that reason, I will use moral non-naturalism as the main target when evaluating debunking arguments. There are four primary reasons for this choice.

First, it will help orient and focus the discussion, and keep it from sprawling in too many directions. Second, it will help keep us focused on the epistemological rather than metaphysical issues involved. As Ramon Das has argued, when debunkers evaluate, say, metaethical naturalist proposals, they often do so by evaluating, not any epistemological claim, but rather various metaphysical claims, such as whether it is plausible that moral facts can be reduced (without reminder) to, identified with, wholly constituted by, or fully grounded in natural facts.[80]

When targeting non-naturalist views, debunkers sometimes *grant* the non-naturalist their metaphysical commitments and then run their argument purely in the epistemological domain. Insofar as debunking arguments are supposed to be epistemological and not metaphysical arguments, they might seem to, on their own, cause problems most directly for the non-naturalist.[81]

Third, it seems reasonable to assume that if non-naturalists can answer the challenge, given that they are granted their decidedly substantial metaphysical claims, the prospects should look very promising for other views as well.[82] Fourth, as mentioned, many debunkers have thought that non-naturalists have few resources with which to reply to or block debunking arguments.[83] If one takes this view, targeting non-naturalists should constitute a low bar to pass for testing the prospects of debunking arguments. If, on the other hand, one thinks that non-naturalism deserves careful attention, it would seem prudent to see how it fares against what is arguably one of the most significant challenges facing the view.

[77] Street (2011); cf. Golub (2017) who discusses a possible reply on behalf of the quasi-realist.

[78] Klenk (2018, 16); Joyce (2016b, 126–27).

[79] Joyce (2016a, 174).

[80] Das (2016).

[81] Joyce (2016a, 157) suggests that a successful articulation of moral naturalism would be a defeater *for the defeater* provided by the evolutionary genealogical information.

[82] *Pace* Bogardus (2016).

[83] Joyce (2006, 210).

Given the above, I believe any drawbacks to exploring the prospects of debunking arguments as faced by a moral non-naturalist are outweighed by the advantages. Furthermore, despite choosing to employ the moral non-naturalist as the debunker's main target, I will often explain how my conclusions relate to other metaethical views as well.

Let me now set out how I understand the commitments of moral non-naturalism. To begin, moral non-naturalism consists of several claims that are not unique to it, but which are accepted by a diverse range of moral realists. These include the claims that

> (1) there are properties or relations corresponding to moral predicates, that (2) moral sentences predicate these properties, and that (3) moral thoughts represent actions, people, or things as having these properties. They think that (4) some of these sentences and thoughts predicating moral properties are true and that (5) we believe the propositions they represent when we accept them.[84]

This makes moral non-naturalists psychological and linguistic *cognitivists* in the sense of taking moral thought and talk to represent the world and the properties in it. I will assume throughout that our moral judgments are best understood as beliefs, and that it is fitting to apply some framework of epistemic justification to them. Beyond this, moral non-naturalism, as I understand it, consists of four further claims. A non-naturalist, as I understand her,

(6) denies that moral facts are reducible (without remainder) to, identical with, wholly constituted by, or fully grounded in natural facts;[85]

(7) holds that the fundamental moral facts are stance-independent, in the sense that they do not obtain in virtue of the attitudes, practices, or customs of human beings;[86]

(8) denies that moral facts (or properties) have causal powers;[87]

[84] van Roojen (2015, 253–54).
[85] Cf. Enoch (2011, 100–105; 2019, 3–4). An exception is perhaps Shafer-Landau, who considers himself a non-naturalist but who subscribes to "the non-identity of moral and descriptive properties, while allowing the moral to be entirely and exhaustively constituted by the descriptive" (2003, 76).
[86] Cf. Enoch (2011, 3–4); Schechter (2018a, 444).
[87] Cf. Enoch (2011, 7); Schechter (2018a, 445). A motivation for this claim is that natural phenomena are never causally acted upon by non-natural phenomena. An exception is Oddie (2005, chap. 1), who claims both to be a non-naturalist and take moral facts to have causal powers. As he does not accept (8), he is not classified as a non-naturalist by my taxonomy.

(9) is not a quietist in the sense of claiming that their moral metaphysics simply does not engage with the question of whether the entities it postulates "really exists," or that our moral discourse leaves no 'ontological footprint.'[88] Hence, the robust non-naturalist moral realist acknowledges making full-blooded ontological claims.[89]

The view outlined here concerns the nature of moral facts as well as our moral thought and talk about them. When I say that certain sets of facts—moral or otherwise—are 'non-natural', I mean that they, *mutatis mutandis*, satisfy (6)-(8). When it comes to the nature of truth and facts itself—the correct metaphysics of truths and facts—I will stay non-committal as far as possible, and I will use these terms interchangeably.[90]

Let me now sum up the ways in which I have narrowed down the scope of the project so far. The following chapters will explore how debunking arguments, including evolutionary debunking arguments, could show that no moral belief is justified or amounts to knowledge. I will primarily consider the threat that such arguments pose to a moral non-naturalist. For this reason, I will also prioritize the replies to debunking arguments that are available to the moral non-naturalist.

1.5 Overview of the Project

The project undertaken in the following pages seeks to evaluate the prospects of global moral debunking arguments. This will be done in two stages, which are encapsulated by the two parts that follow. **Part I**, covering Chapters 2–4 sets out the types of principles that could explain how genealogical information could uniformly undermine moral beliefs. This will be done by taking a detailed look at three paradigmatic global debunking arguments and extracting general principles from them. Having set out how debunking arguments are intended to work in Part I, Part II moves on to the task of evaluating whether they succeed.

[88] Scanlon (2014, 27) defends the first claim and Parfit (2011, 486) the latter. This condition thus rules out so-called quietist versions of non-naturalism. For critical discussion of such views, see McPherson (2011) and Enoch and McPherson (2017).
[89] Cf. Enoch (2011, 7).
[90] This is similar to Enoch's (2011, 5) own commitments in setting out his non-naturalism.

Part II is concerned with evaluating the various principles set out in Part I, by looking at the prospects for debunking arguments built around them. This evaluation will in large part be carried out by focusing on two issues. On the one hand, I will consider strategies for blocking debunking arguments that are available to non-naturalists. We will also look at internal challenges facing such arguments, of which I consider four important ones: the need for plausible first-order epistemological principles, the threat of generalization, the threat of self-defeat, and the necessity of metaepistemological commitments.

Chapters 2–4 investigate how debunking arguments leverage genealogical factors to globally and uniformly undermine moral beliefs. This is done by considering three classical moral debunking arguments, due to Gilbert Harman (Chapter 2), Richard Joyce (Chapter 3), and Sharon Street (Chapter 4). I show that these three arguments fail to properly explain how their genealogical claims lead to their purported epistemological upshots. I then consider a number of more or less implicit candidate epistemological principles that would allow these arguments to generate the conclusion that no moral belief is justified. Among the principles identified are ones concerning ontological parsimony, explanatory dispensability, epistemic insensitivity, unexplained reliability, epistemic coincidences, and explanatory constraints on rational belief.

In Chapters 5–8, I evaluate debunking arguments built on the various candidate epistemological principles identified in Part I, or on refinements of those principles. **Chapter 5** explores the so-called *reliability challenge* for moral beliefs, which holds that they are subject to a problematic form of epistemic coincidence and that their reliability is unexplained. This in turn is taken to undermine these beliefs. An essential component of the reliability of moral beliefs that would need to be explained is the correlation between one’s own moral beliefs and the facts one takes those beliefs to be about. I then distinguish four different models for explaining this correlation—direct explanation, indirect explanation, accidental explanation, and full-blooded Platonism.

Having set out the reliability challenge and the possible ways to answer it, I consider the non-naturalist’s attempt to provide an indirect explanation of the correlation through the third-factor strategy. I argue that the literature on third-factor strategies is plagued by confusion. When correctly understood as the claim that there is a third factor playing a dual explanatory role—i.e., participating in the explanation of both moral beliefs and moral facts—moral non-naturalists cannot successfully employ the third-factor strategy. The reliability challenge is therefore still a live threat to non-naturalists.

In **Chapter 6**, I discuss a different way in which the non-naturalist can attempt to answer the reliability challenge, namely by opting for an accidental explanation of the correlation. This strategy—which I call the *accidental correlation strategy*—accepts that moral facts and our beliefs about them lack a unified explanation. It therefore accepts that moral beliefs are reliable, to the extent that they are, only coincidentally. Even so, this strategy claims that this fact is not sufficient to undermine moral beliefs, because the non-unified explanation it advances can satisfy what is plausibly required for having reliable, justified beliefs and knowledge.

I argue that this reply is not as easily rebutted as the third-factor strategy and that it seems capable of succeeding, at least in principle. I then consider whether debunkers can invoke modal conditions on knowledge or justification in their argument, in order to strengthen the reliability challenge. Such conditions can either be understood as stating necessary conditions for knowledge, or sufficient conditions for epistemic defeat. If successful, this would block the accidental correlation strategy.

I argue that debunking arguments employing modal conditions do not, at present, succeed. This is either because the epistemological principles they rely on are flawed or because the debunker lacks a plausible case for the claim that moral beliefs fail to satisfy them. At present, non-naturalists are therefore able to successfully block the type of debunking arguments we have considered, by accepting the coincidental reliability of our moral beliefs.

A new wave of debunking arguments defending explanatory constraints on justified belief has attempted to block strategies like the accidental correlation strategy. In **Chapter 7**, I provide a novel argument against such debunking arguments, arguing that they face a dual and intertwined threat of overgeneralization and self-defeat. First, I argue that such explanatory constraints threaten to generalize so as to target both empirical and a priori domains. Second, a debunker who relies on such a constraint faces *prima facie* self-defeat since a belief in the constraint plausibly fails to comply with the constraint itself. Avoiding self-defeat will lead a debunker to take on significant metaepistemological commitments, which risks making the resulting debunking argument unattractive.

I show that debunking arguments that ultimately seek to show our moral beliefs to be unjustified (i.e., non-conditional debunking arguments) face the toughest explanatory burdens when faced with the threat of self-defeat. Conditional debunking arguments fare better, but at the cost of having to propose a uniform metanormative account of the moral and the epistemic domains.

Whichever way a debunker goes, constructing a successful debunking argument will have significantly higher explanatory burdens than has hitherto been recognized.

Chapter 8 synthesizes the lessons from the preceding chapters and provides a diagnosis of why debunking arguments face such steep obstacles. I then suggest two conditions of adequacy for any future debunking argument that would allow them to navigate the threats of generalization and self-defeat. I also outline some future directions that debunkers will need to pursue in order to rehabilitate the prospects for a successful global moral debunking argument.

Part I

How Do Debunking Arguments Debunk?

2 Harman's Explanatory Challenge

2.1 Introduction

The aim of this and the next two chapters—i.e., Part I—is two-fold. First, I set out, in some detail, three paradigmatic global debunking arguments targeting moral beliefs. The first, due to Gilbert Harman (this chapter), is something of a precursor to subsequent evolutionary debunking arguments. Harman's argument provides a springboard for setting out two prominent evolutionary debunking arguments—one due to Richard Joyce (Chapter 3), the other to Sharon Street (Chapter 4).

Setting out these three global debunking arguments facilitates the second task of these chapters: to highlight a selection of principles that global moral debunking arguments rely on, either implicitly or explicitly. These are the principles tasked with explaining how genealogical information can constitute sufficient grounds for uniformly undermining moral beliefs. The task of evaluating the plausibility of these principles is one we will pursue in Part II. For ease of reference, I end each chapter in Part I by listing the principles discussed in connection with each argument, as well as providing a road map for where the principles will be discussed and evaluated in Part II.

This chapter begins by setting out Harman's challenge to the explanatory relevance of moral facts (§2.2). It then considers an argument that the challenge fails to provide a relevant model for understanding later evolutionary debunking arguments (§2.3). I claim that this is mistaken, and that Harman's challenge serves up an informative model for understanding the nature of both evolutionary as well as non-evolutionary debunking arguments. I then show that this result is a double-edged sword, as it renders evolutionary debunking arguments subject to many of the objections facing Harman's challenge. I end the chapter by providing a summary and restating the general principle claimed to underlie Harman's challenge (§2.4).

2.2 Explanatory Indispensability

Humans abhor explanatory voids. This fact has resulted in the liberal use (and abuse) of all manner of explanatory models throughout human history—from supernatural agents and anthropomorphized powers to folk psychology. Through such models, we have sought to explain various features of daily life, such as illness and disease, thunder and lightning, and the changing of the seasons. And just like changing seasons, existing explanatory models are superseded by new ones, such as when the attribution of the seasonal cycle to the whims and wills of deities was replaced with a model based on Earth's heliocentric orbit and axial tilt.

It should be uncontroversial that providing such explanations, and in recent centuries, providing scientific explanations, has a central place in human affairs. Some have gone further and claimed that playing an important explanatory role is one of the best, and perhaps only, criteria for determining what we should believe. That is to say, roughly, that we should only believe that p, if p plays some important explanatory role.[1] Call any view that takes explanation to be the primary, or even exclusive, determinant of a belief's justificatory status *explanationism.*

Harman was one of the first to develop a full-fledged epistemological framework that makes explanatory roles central to justified belief, and therefore to justified ontological commitment.[2] This type of explanationist account of justified belief provides an important backdrop for the debunking argument we will look at in this chapter—Harman's explanatory challenge.

Consider Harman's famous—although now somewhat dated—example of a physicist observing a vapor trail in a cloud chamber.[3] The physicist takes the appearance of the vapor trail to indicate that there is a proton in the cloud chamber. Given background beliefs about the operation of such chambers, the physicist more or less automatically makes the judgment that "there goes a proton" upon seeing the vapor trail.

If we entertain a skeptical mindset, we can ask "Why think that there really *is* a proton in the cloud chamber?" or, more generally, "Why think that protons exist?" Harman's answer is that the existence of protons is *indispensable* to

[1] Lycan (2002) presents a taxonomy of views defending some such relation between explanation and justification.

[2] An early formulation of this framework is found in Harman (1965). For a mature statement, see Harman (1986a).

[3] Harman (1977, 6). Cloud chambers went out of use in research contexts in the 1960s and were superseded by other methods of detecting particles. Today they are mostly used for educational and outreach purposes.

the *best explanation* of the physicist's belief that there is a proton in the cloud chamber. Harman himself is concerned specifically with 'observations', in the sense of immediate, non-consciously inferred judgments.[4] However, his central contention generalizes to occurrent beliefs in general—whether immediate or reflective, or consciously or unconsciously inferred.[5]

Let us unpack this generalized version of Harman's claim. What explains the occurrence of the physicist's belief that there is a proton in the cloud chamber? Many factors contribute to the explanation, such as the presence of a vapor trail and the physicist's education and instruction in the use of cloud chambers. In addition, Harman claims that the existence of protons is an ineliminable component of the best explanation of the physicist's belief. After all, if we had to explain the physicist's belief without recourse to protons, that explanation would likely be a worse one. Harman therefore thinks we are justified in believing that protons exist because they play an *indispensable role in the best explanation of some of our beliefs*.

None of the above explains why we should think that being explanatorily indispensable in this way makes some entity (or property) more likely to exist (or be instantiated). Why that should be so is a vexed and contentious issue.[6] Perhaps, as some have suggested, this is simply the point at which requests for further epistemic justification must come a to halt.[7] In any case, the general idea is clear enough. On this picture, indispensability to important explanatory projects is what justifies ontological commitment.

A more precise version of this claim is that an agent gains defeasible justification for believing in whatever is indispensable to the best explanation of some belief that agent holds. When the best explanation of a belief implies the truth of an existence claim, we gain defeasible justification for believing that whatever is so implied exists.

In contrast, if some purported class of facts is not indispensable to the explanation of *any* of our beliefs, they risk being mere explanatory idlers. With respect to such *explanatorily dispensable* facts, Harman thinks, we find ourselves "lacking any evidence" for their existence.[8] When it comes to moral facts, the central question therefore concerns their explanatory role. What do

[4] Harman (1977, 6).

[5] Cf. Joyce (2006, 186). Note that in this context, observations and beliefs are understood as mental events rather than as a propositional content.

[6] See Douven (2011, sec. 3.2) for discussion.

[7] Enoch and Schechter (2008, 563–67).

[8] Harman (1977, 13).

they explain? And more specifically, are they indispensable to the best explanation of some of our beliefs? In an influential ethics textbook, Harman argued that the answer to the latter question appears, at least at first blush, to be ‘no’.[9]

Many have taken Harman’s challenge to be that of explaining how moral facts can enter into causal explanations of beliefs (or of anything else, really).[10] However, if causal explanatory relevance is necessary for justified belief in the existence of a set of facts, then the moral non-naturalist has trivially failed the challenge. The non-naturalist denies any causal powers on behalf of moral facts. On the causal construal of Harman’s challenge, a similar fate will likely befall views in other domains that postulate causally inert entities, facts, and properties. Fortunately for the non-naturalist, a straightforward causal constraint on knowledge or justification is implausible and rarely defended.

In fact, the progenitor of the causal account of knowledge, Alvin Goldman, explicitly excluded that account from applying to a priori knowledge.[11] The account is even known to run into trouble with the a posteriori domain. A simple causal constraint on knowledge, such as that in order to know that p, it is necessary that one’s belief that p be caused by the fact that p, would block knowledge of the future. Barring backwards causation, facts about the future can never cause our present beliefs.

It might be possible to circumvent such issues, as we will discuss in Chapter 7. Even so, we can allow that whatever Harman’s challenge was originally intended to be, it can be reconceived as requiring indispensability for *some* type of explanation of our beliefs, though that explanation need not be causal.

Harman illustrates his challenge with a scenario that is supposed to contrast with the physicist studying the cloud chamber. In this second scenario, a group of children dose a cat in gasoline and light it on fire.[12] A nearby onlooker, who watches the event unfold, judges that what the kids are doing is wrong. The question is then whether this case is parallel to the cloud chamber scenario. For it to be parallel, some moral fact(s) must be indispensable to the best explanation of the onlooker’s belief that what the kids are doing is wrong.

To determine this, one first needs to determine what constitutes the best explanation of the onlooker’s belief. For a start, that explanation should presumably invoke the moral psychology and outlook of the onlooker. Once those factors are invoked, however, Harman believes we already have a sufficient explanation of the belief, and that it is simply superfluous to make any further

[9] Harman (1977). See also his (1986b).

[10] Shafer-Landau (2003, 99–100).

[11] Goldman (1967, 357).

[12] Harman (1977, 4).

assumptions about the existence of moral facts (or about the instantiation of moral properties). As he recaps his point in a later text:

> [O]ur having the moral beliefs we have can be explained entirely in terms of our upbringing and our psychology, without any appeal to an independent realm of values and obligations.[13]

Harman does not elaborate at any length on how one's upbringing and psychology are meant to explain the generation of one's moral judgments, but finds it plausible that it comes down to social conventions.[14] We could alternatively fill in that story in whatever way we desire, as long as it does not make an indispensable reference to moral facts. After having filled in such a story, Harman thinks, we would have a fully satisfactory explanation of why the onlooker believes that the act of setting the cat on fire is wrong. There would then not be "any obvious reason to assume anything about 'moral facts,' such as that it really is wrong to set the cat on fire."[15]

Even if one *did* postulate the existence of moral facts, it can be hard to see how it could contribute anything of explanatory relevance or otherwise improve on the explanation of the onlooker's belief. This is especially true if the facts in question are non-natural moral facts, as the moral non-naturalist will have it. Such facts cannot, by their very nature, provide any causal explanatory relevance, and it is not clear what other form of explanatory relevance they could have either.

A way of making Harman's challenge more precise is generated by focusing on the tight connection that Harman assumes to hold between explanatory indispensability and justified belief—i.e., his explanationist commitments. Harman is not explicit, in this particular context, about what exactly that relation is. For instance, he does not say whether he holds being explanatorily indispensable to be a necessary (and perhaps also sufficient) condition for having a justified belief, or for justified ontological commitment.[16] Harman nonetheless does seem, at least implicitly, to assume that explanatory indispensability *is* necessary.[17] This allows for an interpretation of Harman where his challenge relies on a principle such as the following.

[13] Harman (1984, 32).

[14] Harman and Thompson (1996, 26–27).

[15] Harman (1977, 7).

[16] Elsewhere, outside of his challenge to moral facts, Harman is explicit about this, and a refined version of his view on the relation between justification and explanation is found in his (1986a).

[17] As many have pointed out, this claim, and the principle itself, will have to be amended in various ways in order to be plausible (Pust 2001, 231; Schechter 2018a, 445–50). Such details can be overlooked for our purposes here.

> EXPLANATORY REQUIREMENT
>
> An agent's belief in a fact F (or the instantiation of a property P) is justified only if F (or P) is indispensable to the best explanation of some belief held by that agent.[18]

This principle captures the idea that the best guide we have to reality is one that requires that any putative facts (or properties) earn their keep by explaining why we believe what we do. If you believe that p, but the fact that p plays *no part whatsoever*, or only a dispensable part, in explaining that belief (or any other belief of yours), then something has gone wrong, epistemically speaking.[19] With such a requirement in hand, Harman's challenge consists in the question of whether our belief in the existence of moral facts can satisfy his EXPLANATORY REQUIREMENT.

On this construal, Harman's challenge can concisely be stated as follows.

> HARMAN'S CHALLENGE
>
> (1) Belief in the existence of moral facts is justified only if moral facts are indispensable to the best explanation of some of our beliefs.
>
> (2) Moral facts are never indispensable to the best explanation of any of our beliefs.
>
> (3) Therefore, belief in the existence of moral facts is unjustified.

Premise (1) is supported by Harman's EXPLANATORY REQUIREMENT. Premise (2) is supported by the generalization of Harman's scenario involving the onlooker witnessing a cat being set on fire. (Or by the fact that the non-naturalist might be thought to lack *any* plausible suggestion for how moral facts could possibly be explanatorily relevant with respect to our moral beliefs).

Having set out Harman's challenge, I will now briefly consider some objections to it. My intention is not to defend or argue against the challenge, but

[18] Cf. Harman (1977, 13). For alternative formulations of such a principle, see Wright (1992, 177); Pust (2001, 231); Cuneo (2003, 351).

[19] One might also wish to allow for the possibility that one's beliefs about p somehow constitutes or constructs the fact that p. We will explore explanationist principles that allow for this in Chapter 7.

merely to set it out as a backdrop against which to interpret subsequent evolutionary debunking arguments. As we shall see, such arguments share with it significant similarities and presuppositions.

There are two main ways in which to respond to Harman's challenge—either by trying to answer it, or by rejecting it. By taking the first line of reply, one accepts premise (1), and thereby the EXPLANATORY REQUIREMENT, and goes on to argue that moral facts actually *do* play the required explanatory role. This would involve showing that moral facts, despite appearances, are indispensable to the best explanation of some of our beliefs.

One such strategy consists in showing moral facts to be explanatorily indispensable in virtue of their relation to natural facts.[20] For instance, this could be done if moral facts were reduced (without reminder) to, identified with, wholly constituted by, or fully grounded in natural facts.[21] Thus, securing a sufficiently strong connection between moral and natural facts could perhaps answer Harman's challenge. Harman himself ultimately believes the challenge can be met in this way, as he argues that moral facts can be reduced, in a broad sense, to certain natural facts.[22]

Another strategy is to argue that we are capable of "moral perception," such that we stand in a relation to moral facts that is relevantly similar to how we relate to ordinary physical objects through visual perception. If defensible, the onlooker in Harman's case could then be claimed to "perceive" the wrongness of the kids' action more or less directly. Such moral perceptual contact with moral facts (or properties), in turn, could then be held to be indispensable to the explanation of the onlooker's belief.[23]

The second line of reply—rejecting the challenge—rejects the EXPLANATORY REQUIREMENT, and thereby premise (1). This strategy can take many forms. It could be argued that the EXPLANATORY REQUIREMENT itself—that very epistemic fact—seems like an unlikely candidate when it comes to being indispensable for the explanation of any of our beliefs. Hence, the principle appears to be self-defeating.[24] Alternatively, it could be argued that not even

[20] Harman (1977, 13).

[21] For some examples of arguments that moral facts are capable of playing the required explanatory role in virtue of their relation to natural facts, see Sturgeon (1985; 2006), Railton (1986), Boyd (1988), Brink (1989, 182–97), and Copp (1990).

[22] Harman claims that "there is empirical evidence that there are (relational) moral facts" (1977, 132). One might of course doubt that even such relations make things all that much easier for the naturalist, but as I intend to focus on non-naturalism and its strategies, I will set such issues aside.

[23] See Werner (2020) for an overview of the literature on moral perception. For a relevant exchange, see Faraci (2015) and Werner (2018).

[24] Pust (2001).

empirical facts, like the existence of protons, pass Harman's test and that this shows it to be mistaken.[25]

This line of reply could also hold that the kind of explanatory role that is required for justified ontological commitment is different from what Harman suggests. Some have argued that while moral facts do not participate in the best explanation of our *beliefs*, they are nonetheless indispensable for the best explanation of moral, aesthetic, or normative facts.[26]

Another way of developing this reply is by holding that while being indispensable to the best explanation of our beliefs is sufficient for generating defeasibly justified ontological commitment, it is not a necessary condition for it. There might be other ways to justify a belief in the existence of some fact than by an appeal to its indispensability for explanatory purposes. Enoch has argued that non-explanatory forms of indispensability might equally suffice.[27]

Lastly, one might reject the broader epistemological assumptions of Harman's challenge—it is explanationist commitments—and subscribe to some epistemological view of justified belief which does not require any intimate connection to explanation. Consider phenomenological conservatism.

> PHENOMENAL CONSERVATISM
> If it seems to S that p, then, in the absence of defeaters, S thereby has at least some degree of justification for believing that p.[28]

This principle would allow any undefeated, seemingly true belief, whether about morality or some other domain, to be defeasibly justified.[29] If phenomenal conservatism is true, it is possible to have defeasibly justified moral beliefs where those beliefs are not justified by playing some indispensable explanatory role.

More generally, it is not self-evident what rationale there is for claiming that the EXPLANATORY REQUIREMENT is a necessary condition on justified ontological commitment. While it may have some intuitive appeal, defending that claim would require far more work than is usually expected by Harman's

[25] Wright (1992, 190–91).

[26] Moral facts: Schafer-Landau (2003, 104; 2007, 322–32); Sober (2015, 266–67). Aesthetic facts: Gaut (2007). Normative facts: Chang (2004). Cf. Harman (1977, 8–9).

[27] Enoch (2011, chap. 3).

[28] Huemer (2007, 30). For defense of the principle see Huemer (2001; 2007; 2014). For criticism, see Siegel (2012), White (2006), and Steup (2013).

[29] This particular principle will likely only be compatible with certain internalist views of epistemic justification.

challenge, as one would need to argue for the requisite comprehensive epistemological view, and against competing views, such as phenomenal conservativism.[30]

Even those sympathetic to Harman's challenge tend to modify it in various ways. Some suggest eliminating the reliance on the concepts of 'best explanation' and/or 'indispensability'.[31] Others, that the challenge should not be framed in terms of necessary conditions for justified ontological commitment.[32] Instead of setting out a necessary condition for defeasibly justified belief, on this view, it should be seen as a constraint on justified belief. That is, the explanatory condition should be understood as a sufficient condition for epistemic defeat. This would allow it to be a threat even to the adoptee of non-explanationist views of justification, such as phenomenal conservativism, as it would then count as a defeater. I will discuss a spiritual successor of Harman's challenge which incorporates such elements in Chapter 7.

As I do not intend to defend Harman's challenge as set out above, I will not dwell further on these issues here. My intention has been to set out the challenge—and objections to it—as a backdrop against which to interpret subsequent evolutionary debunking arguments. Among these is the issue of what explanatory role, if any, moral facts would need to play in order to be vindicated, and whether they in fact do play such a role.

2.3 From Harman's Challenge to Evolutionary Debunking Arguments

Harman's challenge, as formulated above, concerns the epistemic import of moral facts being, at best, explanatorily redundant with respect to our beliefs. If explanatory indispensability is a necessary condition on justified ontological commitment, as in the EXPLANATORY REQUIREMENT, then that import is that a belief in the existence of moral facts is unjustified. Harman's challenge is therefore an argument where, by considering the genealogical features of moral beliefs, we can come to conclude that they are uniformly unjustified.[33] This makes Harman's challenge a global moral debunking argument, as I defined them in §1.3.

[30] Cf. Shafer-Landau (2007, 320–22); Braddock (2017, 93).

[31] Wright (1992, 189–91); Majors (2007, 5).

[32] Schechter (2018a, 447).

[33] Harman (e.g. 1977, 12) sometimes claims that it might incline us towards some non-cognitivist view as well.

It is worth noting that Harman's challenge concerns the explanation of how *individuals* form *token* beliefs. Harman himself appeals to social conventions in order to explain the formation of individuals' moral psychology and its generation of token moral judgments.[34] In the time since Harman formulated his challenge, the evolutionary roots of human morality have received extensive attention and research. This has led to attempts at constructing a plausible, empirical evolutionary genealogy for human moral psychology.

Many have argued that Harman's challenge, while not itself concerned with evolutionary genealogies, provides a particularly relevant framework for understanding subsequent global moral debunking arguments. Such arguments seek to explain human moral psychology by recourse to our evolutionary history.[35] In this way, one might think that Harman's challenge can be buttressed by findings and speculation concerning the evolutionary origins of human morality.

Many have assumed that Haman's challenge is highly relevant for the formulation of evolutionary moral debunking arguments. Andreas Mogensen has argued that this assumption is at best misguided and at worst a result of fallacious reasoning.[36] Mogensen argues that while Harman's challenge is concerned with the individual-level explanation of the adoption of token moral beliefs, evolutionary debunking arguments are concerned with a different type of explanation altogether. Evolutionary debunking arguments deal in explanations of population-level distributions of moral judgments, which seek to explain why a trait has been selected for in a population over time. This type of explanation invokes *ultimate causes*—the causes of population-level distributions of features in a population over time.[37]

When explaining some feature of an organism, explanations invoking such ultimate causes are fully compatible with explanations operating at the level of an individual that concern the *proximate causes* of that feature. Although compatible, such explanations can be different in kind. These separate, but compatible levels of explanation can cause confusion, as Mogensen claims have happened in the literature on evolutionary debunking.

In an analogy from biology, Mogensen uses the example of a species of insects that have a particular color pattern that serves as camouflage by allowing the insects to blend in with the surrounding environment.[38] The ultimate

[34] For Harman's view, see Harman and Thomson (1996, 26–27).

[35] Joyce (2006, 184 ff); Clarke-Doane (2016).

[36] Mogensen (2015).

[37] The proximate/ultimate distinction employed by Mogensen stems from Mayr (1961).

[38] Mogensen (2015, 198).

cause of this trait being selected for is that it helps individuals possessing it to avoid predators and therefore confers a relative fitness advantage. The proximate cause of the coloration pattern for any individual would be of a different kind altogether; the trait could arise in an individual because of diet or stem from a random mutation.

Even if the ultimate causes of moral beliefs—conferring a relative fitness advantage—are shown to render moral facts explanatorily redundant, that leaves it open that moral facts are nonetheless indispensable to the explanation of the *proximate causes* of moral beliefs. Mogensen quotes Robert Nozick as having recognized the danger of conflating these types of explanations.

> If ethical behavior increases inclusive fitness, this will explain the spread of such behavior in the population. Yet each individual's behavior, ancestor or descendant, might be explained by her recognizing certain ethical truths and acting on them.[39]

Mogensen claims that proponents of evolutionary debunking arguments have failed to appreciate the different levels of explanation involved. To see the problem clearly, consider that Harman's challenge concerns whether moral facts are ever indispensable to the best proximate explanation of an individual's beliefs.

Either they are, or they are not. If moral facts *are* indispensable to such explanations, then it is not clear that we should be all that worried by evolutionary debunking arguments aiming to establish something about their dispensability for ultimate explanations. Whatever such arguments establish, individuals' beliefs would nonetheless satisfy the EXPLANATORY REQUIREMENT (at least when it is disambiguated to concern proximate explanations).

If, on the other hand, moral facts *are not* indispensable for the best proximate explanation of any moral (or non-moral) judgment, the claim that *neither* are they part of the best ultimate explanation of our moral beliefs might not seem to add much argumentative force.

Mogensen's conclusion is that insofar as evolutionary debunkers leave out any discussion of the proximate causes for moral beliefs, their debunking argument straightforwardly fails. Mogensen claims that Joyce's argument, which we will consider in the next chapter, fails in exactly this way.[40] Conversely, if a debunker were to defend the argument satisfactorily at the proximate level—i.e. Harman's challenge—tacking on a similar and further claim

[39] Nozick (1981, 345), quoted in Mogensen (2015, 197–98).
[40] Joyce (2006).

about ultimate causes would be of relatively minor importance. Hence, when taken on their own, evolutionary debunking arguments either fail or are uninteresting.

An alternative view of the relation between Harman's challenge and evolutionary debunking arguments has been suggested by William J. FitzPatrick.[41] He claims that evolutionary debunking arguments, and their proponents, do not make the mistake of being exclusively concerned with the ultimate level of explanation. Instead, any spelled-out evolutionary debunking argument will include a theory of proximate-level explanations of moral belief, just like any fully spelled-out evolutionary story about the color patterns of some insect will involve proximate-level explanations of that trait. Such proximate explanations will be a natural complement to the population-level story, and without both, the theory would be missing an essential component.

While Mogensen's point about being mindful of the different types of explanation is well taken, the more charitable reading of the relationship between Harman and subsequent evolutionary debunkers allows for the development of the type of combined ultimate/proximate story suggested by FitzPatrick. However, being forced to co-opt Harman's challenge is both a blessing and a curse for the evolutionary debunker. By doing so, debunkers will invariably need to wade into the issues generated by Harman's challenge.

We saw some of these issues—concerning explanatory relevance, best explanations, and theories of epistemic justification—outlined above. When running an evolutionary debunking argument, these concerns regarding explanations at the proximate level therefore need to be added *on top of* whatever issues are raised by running the argument at the level of ultimate explanations. Evolutionary debunking arguments will therefore be even more demanding to defend than merely defending a version of Harman's challenge.

With these clarification in mind, I will move on to set out two classical formulations of evolutionary debunking arguments in the next two chapters.

2.4 Summary & Preview

In this chapter, we have seen how Harman's challenge can be understood as a global moral debunking argument. It plausibly assumes that certain types of explanatory connections are required for justified belief. In particular, it requires that any purported facts must be indispensable for the best explanation

[41] FitzPatrick (2016).

of some belief. This was encapsulated by the following necessary condition on justified ontological commitment.

> EXPLANATORY REQUIREMENT
>
> An agent's belief in a fact F (or the instantiation of a property P) is justified only if F (or P) is indispensable to the best explanation of some belief held by that agent.

Harman's challenge, with its focus on explanatory connections, has been a formative influence for many evolutionary debunking arguments. Joyce's debunking argument, discussed in the next chapter, is explicitly modeled on Haman's challenge, and Joyce expands it in order to show that moral facts are explanatorily indispensable for explaining the selection pressures affecting moral beliefs.

In Chapter 7, we will explore a debunking argument that closely resembles Harman's challenge, and which involves the defense of an explanatory constraint on justified belief. The challenge there is motivated by the same explanationist commitments as are embraced by Harman, although refined in an attempt to avoid the objections leveled against him.

3 Joyce's Evolutionary Debunking Argument

3.1 Introduction

In this chapter, we will explore how Richard Joyce's evolutionary debunking argument attempts to employ genealogical information in order to uniformly undermine moral beliefs. Joyce's argument is explicitly modeled on Harman's challenge, so it will allow us to evaluate the continuity from non-evolutionary debunking arguments to their evolutionary counterparts.

I begin by setting out Joyce's debunking argument, both in relation to his metaethical error theory and as a stand-alone epistemological argument (§3.2). I then consider a number of epistemological principles that are either refinements of principles explicitly suggested by Joyce, or which can otherwise reasonably be attributed to him (§3.3). This is somewhat complicated by the fact that Joyce has presented many versions of his argument over the years, where those versions have relied on somewhat different grounds. I consider principles having to do with parsimony, explanatory dispensability, the modal profile of moral beliefs, and explanatory constraints. I end by providing a summary of the chapter and restating the principles claimed to underlie Joyce's debunking argument (§3.4).

3.2 Joyce's Debunking Argument(s)

Joyce has defended evolutionary debunking arguments in a number of publications over the last two decades.[1] In his earlier work, he employs a debunking argument as a supplementary aid alongside more traditional metaethical arguments in the context of arguing for a moral error theory.[2] The moral error theory Joyce defends holds that moral claims uniformly fail to be true. The phrase 'fails to be true' signifies that Joyce defends a version of the error theory where moral discourse is claimed to consist in assertions that are neither true

[1] Joyce (2001; 2006; 2016a, chap. 7; 2016b).
[2] Joyce (2001, chap. 6).

nor false.[3] Such a radical view clashes with strongly held intuitions and incurs explanatory burdens that Joyce believes a debunking argument can help alleviate.

> A proponent of an error theory—especially when the error is being attributed to a common, familiar way of talking—owes us an account of why we have been led to commit such a fundamental, systematic mistake. In the case of morality, I believe, the answer is simple: natural selection.[4]

Joyce believes that an evolutionary debunking argument can do more than fill in explanatory gaps; he thinks it could be sufficient, independently of his arguments for the error theory, to establish that moral beliefs are uniformly unjustified. In later work, Joyce has developed such a stand-alone global moral evolutionary debunking argument.[5] When Joyce runs the argument in this fashion—independently from his error theory—it is compatible with a form of skeptical moral realism that his error theory is not. Even if our moral beliefs turn out to be epistemically unjustified, that does not rule out the metaphysical possibility that there are moral facts.

In addition to separating the epistemic debunking argument from his error theory, Joyce has also come to present the argument in a more specific and hedged form. In its most developed version, he makes the argument conditional on a particular hypothesis about the innateness of human morality. His considered debunking argument, therefore, is that *if* morality is innate, *then* our moral beliefs are undermined.[6] Joyce's support for this conditional argument consists, first, of what he takes to be a substantial, but not conclusive, defense of the hypothesis that morality is innate. Second, it depends on an argument to the effect that the innateness of morality would undermine moral beliefs. I will discuss these two claims in order.

The innateness hypothesis championed by Joyce is that "human morality is innate."[7] By 'morality', Joyce has in mind our *capacity* and *tendency* to make moral judgments, where such judgments are identified by their possession of a sufficient number of features among a cluster of paradigmatic attributes.[8] Joyce does not take a stance on exactly which features are necessary or jointly sufficient for a judgment to qualify as moral. Instead, he lists attributes that

[3] Joyce (2001, 6–9).

[4] Joyce (2001, 135).

[5] Joyce (2006; 2016a, chap. 7).

[6] Joyce (2006, 1–2); Joyce (2016a, chap. 8).

[7] Joyce (2008, 213).

[8] Joyce (2006, 70–71).

typically characterize them, including the expression of conative attitudes, purporting to be inescapable, being concerned with interpersonal relations, involving notions of desert and justice, etc.[9]

Joyce is explicit that he intends to separate out morality from related but distinct phenomena such as pro-sociality and altruism. Showing that a mechanism for generating pro-social or altruistic behavior is innate would therefore not be sufficient to show that the same is true of morality.

The hypothesis that morality is innate in humans, in this sense, does not entail that there is such a thing as *innate moral beliefs*, in the sense of a particular belief content being innate. Rather, it is the claim that humans have an innate mechanism that affords us the ability to "categorize the world in morally normative terms; moral *concepts* may be innate even if moral beliefs are not."[10] While not entailing it, the innateness hypothesis is nonetheless compatible with the claim that certain moral beliefs are innate.

A trait being innate, in Joyce's usage, means that "the present-day existence of the trait is to be explained by reference to a genotype having granted ancestors reproductive advantage, rather than by reference to psychological processes of acquisition."[11] As for our capacity and tendency to make moral judgments, this Joyce thinks has been "biologically selected for in our lineage."[12]

As for why we should think that our capacity and tendency to make moral judgments have been selected for in our own lineage, Joyce relies on findings in evolutionary biology that show that (non-moral) helping behavior has evolved because of its fitness-enhancing effects.[13] Given this, Joyce extrapolates that an innate tendency to make moralized judgments about such helping behavior would similarly enhance reproductive fitness by constituting a "bulwark against the temptations of short-term profit."[14]

Making moralized judgments could help agents overcome variability or weakness in motivation that might otherwise have prevented the occurrence of helping behavior or even caused anti-social behavior that would ultimately have been fitness-reducing. For moralized judgments to have this bulwark effect, they would have to be able to block the prudential analysis that agents engage in with respect to non-moral actions. In this way, agents might perform

[9] Joyce (2006, 70–71).

[10] Joyce (2006, 181), italics in the original.

[11] Joyce (2006, 2).

[12] Joyce (2006, 140). For the distinction between selection for/of, see Sober (1984).

[13] E.g. Hamilton (1964a; 1964b).

[14] Joyce (2001, 213).

helping behaviors rather than trying to free-ride or exploit opportunities, despite the latter being recommended by a purely prudential analysis.

So much for Joyce's tentative defense of the hypothesis that human morality is innate. Why think that this hypothesis, if true, has the power to uniformly undermine moral beliefs? To pump our intuitions about the epistemological upshots of the innateness hypothesis, Joyce employs what he intends as a relevantly analogous scenario. The scenario involves taking a pill—a 'Napoleon pill'—that makes you form beliefs at random about Napoleon (whereas you would otherwise not have formed any beliefs at all about Napoleon).

> Suppose […] that you discover beyond any doubt that you were slipped one of these pills a few years ago. Does this undermine all the beliefs you have concerning Napoleon? Of course it does.[15]

Joyce goes on to make explicit the parallel to an innate tendency to make moral judgments.

> Instead of Napoleon beliefs suppose it is moral beliefs, and instead of belief pills suppose it is natural selection. Were it not for a certain social ancestry affecting our biology, the argument goes, we wouldn't have concepts like obligation, virtue, property, desert, and fairness at all. If the analogy is reasonable, therefore, it would appear that once we become aware of this genealogy of morals we should (epistemically) […] cultivate agnosticism regarding all positive beliefs involving these concepts.[16]

The worry, then, is that we have an explanation—a genealogy—of how the capacity and tendency to make moral judgments became widespread in the human population, and that this genealogy undermines the relevant beliefs. Joyce is not explicit about how the availability of such an evolutionary genealogy succeeds in undermining moral beliefs, and he relies to a large extent on analogical reasoning from the Napoleon case. It is therefore not entirely clear whether Joyce conceives of his debunking argument as relying on some particular or general principle or whether it merely relies on an analogy with an intuitive case.[17]

It is clear enough that Joyce's central claim is that a fully satisfactory explanation of our capacity and tendency to produce moral judgments requires only the claim that possessing such concepts and making such judgments have

[15] Joyce (2006, 181).
[16] Joyce (2006, 181).
[17] Cf. Wielenberg (2016, 505).

proved evolutionarily beneficial by being fitness-enhancing through fostering cooperation and helping behavior.

Recalling the proximate-ultimate distinction, we should understand the evolutionary story Joyce provides as intended to provide a unified account of explanatory dispensability with respect to moral facts. Not only are moral facts dispensable when it comes to the proximate causes of why any individual forms a given moral belief, but they are equally dispensable for the best explanation of why a certain moral-cognitive architecture has become widespread in the human population over time.[18] Like Harman, Joyce thinks such explanatory dispensability undermines epistemic justification.

> My contention, then, is that moral nativism can have epistemological implications […] In particular, any epistemological benefit-of-the-doubt that might have been extended to moral beliefs […] will be neutralized by the availability of an empirically confirmed genealogy of those beliefs that nowhere implies or presupposes their truth.[19]

A pressing question becomes how, exactly, "the availability of an empirically confirmed genealogy of [moral] beliefs that nowhere implies or presupposes their truth" undermines their justificatory status. We will consider this issue in the next section.

3.3 Candidate Principles

In earlier formulations of his debunking argument, Joyce is not explicit about exactly how establishing a genealogy of moral belief that "nowhere implies or presupposes their truth" succeeds in undermining them. One might hold that it seems intuitively obvious that the relevant type of explanatory dispensability undermines. It is more likely, however, that there is a general principle in play. The question therefore becomes what principle(s) underlie Joyce's debunking project. We will return to this below. Before moving on to that issue, it is worth orienting Joyce's notion of 'undermining' to that found in Harman's challenge. We saw that Harman should plausibly be interpreted as taking his Explanatory Requirement to be a necessary condition on justification. That in turn, made his challenge vulnerable to simple denials of the condition, for instance by subscribing to alternative theories of what justifies belief.

[18] Joyce (2006, 123–33) provides some possible mechanisms for the explanation of the proximate causes of moral belief, such as emotional dispositions.
[19] Joyce (2006, 209–10).

In his own discussion of how his arguments might be affected by various epistemological theories of justification, as well as in the quote above, it seems clear that Joyce is comfortable with the idea of allowing that moral beliefs can be antecedently defeasibly justified. Joyce's debunking project should therefore be understood as providing some feature of moral beliefs that undermines any such defeasible justification. Throughout his discussion, Joyce brings up a number of more or less related features that could explain why moral beliefs are uniformly undermined.[20]

3.3.1 Ockham's Razor

Joyce explicitly operates with much of the same theoretical machinery as Harman. Joyce suggests that an operative principle in Harman's challenge is Ockham's razor, and that this principle can do the work for evolutionary debunking arguments as well.[21] Joyce himself does not set out any specific formulation of that principle, but he claims that Ockham's razor dictates that if the existence of moral facts constitutes a mere explanatory idler on top of the natural genealogy of our moral beliefs, then an explanation that omits them should be preferred to one that includes them, other things being equal.[22] This interpretation of Ockham's razor seems to fit with the disambiguation of Ockham's' razor that E.C. Barnes calls the *Anti-Superfluity Principle*.[23]

> OCKHAM'S RAZOR (*Anti-Superfluity Principle*)
> Avoid positing theoretical components which are not required for the purpose of explaining the relevant data.[24]

If non-naturalists accept that beliefs are the relevant data and that moral facts play no indispensable role in explaining that data, they might seem to violate

[20] A feature taken up by Joyce in passing that I will not discuss here is defeaters based on disagreement (Joyce 2006, 216). We return to debunking arguments from disagreement in §8.2.3.
[21] Other debunkers who similarly appeal to Ockham's razor in a debunking context are Ruse (1986, 256) and Ruse and Wilson (1986, 186–87).
[22] Joyce (2006, 189, 195, 210).
[23] Barnes (2000, 369). For an enlightening discussion of debunking arguments relying on Ockham's razor in light of this distinction, see Mogensen (2014, 99–112).
[24] Cf. Barnes (2000, 354). Barnes separates this interpretation of Ockham's razor from its perhaps more famous sibling, which he terms the *Anti-Quantity Principle*. That latter principle "urges that theorists posit as few theoretical components as possible in the construction of explanations of phenomena" (Barnes 2000, 354).

the anti-superfluity principle. In this way, Joyce takes Ockham's razor, on its own, to be able to undermine moral non-naturalism.[25] This is because the non-naturalist, unlike the naturalist, is likely to grant that moral facts do not play a non-superfluous role with respect to explaining our beliefs, moral or otherwise.[26] This, in turn, is because the non-naturalist is committed to moral facts being non-causal.

Despite not being considered by Harman or Joyce, there is a possibility for saving the explanatory relevance of moral facts while denying them causal influence. This strategy could hold that moral facts are indispensable for a *non-causal* explanation of our moral beliefs. Consider for instance traditional epistemic rationalism, which is the view that we have a cognitive faculty or ability—rational intuition—which affords us a non-empirical, non-causal, quasi-perceptual connection with a priori truths.[27]

On this type of view, moral facts themselves would not be part of the *causal* explanation of why we believe them, as the relevant set of truths would be causally inert. It would nonetheless be the case that we believe some moral truth *because* they are true, and that the truths could therefore be indispensable for the explanation of our belief in them. While such traditional forms of rationalism are thereby capable of securing the required form of non-causal explanatory connection, the view is widely thought to be untenable. In a discussion of competing views of a priori knowledge, Carrie Jenkins states:

> The consensus now […] is that such appeals are unacceptable: we must respect the fact that there is no scientific evidence for any special faculty of the kind envisaged. We can only postulate one as a last resort, if all other accounts of a priori knowledge demonstrably fail to do what is required of them.[28]

Having set aside the possibility of rational intuition, as traditionally conceived, the non-naturalist again faces the threat of Ockham's razor. If the non-naturalist both grants that moral facts are explanatorily superfluous and accepts

[25] Joyce does not think this principle is sufficient to rule out moral naturalism, since, if true, such views could conceivably hold moral facts to be explanatorily indispensable.

[26] While Joyce and I both attribute this commitment to the non-naturalist, some proclaimed non-naturalists do not accept it. FitzPatrick (2015, 894 fn. 14) claims that (non-naturalist) realism is doomed if the superfluity claim is accepted. A non-naturalist realist like FitzPatrick will then be faced with explaining exactly *how* non-natural moral facts could be an indispensable part of the explanation of our moral beliefs. We will discuss this challenge repeatedly below.

[27] The view, applied to mathematical knowledge, is attributed to Gödel (1964, 271–72): "[D]espite their remoteness from sense experience, we do have something like a perception also of the objects of set theory, as is seen from the fact that the axioms force themselves upon us as being true." BonJour (1998) defends a similar account of a priori knowledge.

[28] Jenkins (2008, 56).

the *Anti-Superfluity Principle*, they seem to straightforwardly be defeated by it. The first answer to how evolutionary debunking arguments debunk is therefore that the undermining force of genealogical explanations of moral beliefs consists, quite simply, in an appeal to Ockham's razor (at both the level of ultimate as well as proximate causes).

The appeal to Ockham's razor in order to undermine moral beliefs, as construed by a non-naturalist, is too swift, however, and for at least two reasons. The first is that, while Harman and Joyce are singularly concerned with whether moral facts can explain our beliefs, there are plausibly other worthwhile *explananda* as well.[29] For instance, there is a perfectly ordinary sense in which moral facts *do* play an indispensable explanatory role. A certain type of moral fact—moral principles—is arguably required in order to explain particular moral facts.[30] The principle that causing unnecessary pain is wrong (or some such) might plausibly be thought indispensable to the explanation of why pouring gasoline on a cat and setting it on fire is morally wrong.

Moral facts have also been thought to help explain things outside the moral domain. Some have argued that the aesthetic qualities of artworks can in part be explained by appeal to moral considerations.[31] Others that facts about what is to be done, all things considered, are in part explained by moral facts.[32] Relative to *such* explanations, moral facts are far from superfluous.[33]

It might seem question-begging for a moral non-naturalist (or any other type of moral realist) to wheel in moral facts as *explananda*, given that it is the justificatory status of precisely this class of facts that is under scrutiny. It might seem even worse if this is done for the purpose of licensing the explanatory prowess *of moral facts*. We will discuss this type of circularity objection at length in §5.5. There, I will argue that such an objection cannot ultimately be sustained by a debunker. If that is so, then it is not clear what blocks the non-naturalist's claim that moral facts *are* explanatorily indispensable.

This brings us to the second obstacle to employing Ockham's razor as Joyce does. Ockham's' razor—in the form of the *Anti-Superfluity Principle*

[29] Wright (1992, 196–97) argues that explaining intentional states such as beliefs is perhaps the *least* important explanatory role to consider.

[30] Harman (1977, 8–9) was aware of this but thought it irrelevant given his choice of privileged explanatory role. See the discussion in §5.6 below for discussion of the role of moral principles in moral explanations.

[31] Gaut (2007).

[32] Chang (2004).

[33] For claims that normative explanatory tasks such as these are the proper purview of moral facts, see Schafer-Landau (2007, 322–32) and Sober (2015, 266–67).

(as well as in the form of its sibling, the *Anti-Quantity Principle*)—has certain restrictions in its application.[34] According to Barnes, such principles

> are intended to apply to the evaluation of theories insofar as theories are supported by a body of data they purport to explain. Both principles are clearly not intended to apply to other cases of evidential support.[35]

If non-natural moral facts are assumed to be in the business of explaining a body of data—either the psychological states of individuals or the population-level distribution thereof—an appeal to Ockham's razor is unproblematic. However, it is not evident that this is the appropriate explanatory model to saddle moral facts with, especially as conceived by a moral non-naturalist. While it is sometimes the case that we use a body of data to construct moral theories, it can be doubted whether this is the primary mode of ethical theorizing.

As we saw above, one could hold that moral facts are in the business of explaining other moral facts, aesthetic qualities, or normative facts. In that case, Ockham's razor would not apply. More generally, many have thought that moral theorizing is not primarily a scientific or empirical enterprise where we attempt to support our theories by appealing to a body of data—whether human beliefs, behavior or something else. Elliot Sober, in a discussion of why he believes Ockham's razor does not threaten the justificatory status of moral facts, has put forth a similar claim.

> [N]ormative ethical propositions should not be evaluated by their ability to explain descriptive propositions about human thought and behavior.[36]

To rebut this line of reply, the debunker would need to defend the claim that when it comes to justified ontological commitment, *only* theories that aim to explain—and be supported by—empirical data are to be accepted. Such an empiricism could certainly be adopted by a debunker.[37] Were the debunker to adopt it, however, it would make any debunking argument against the non-naturalist entirely superfluous, as a non-naturalist position would be ruled out by *fiat* instead of by any debunking argument appealing to Ockham's razor.

[34] For the *Anti-Quantity Principle*, see fn. 24 above.

[35] Barnes (2000, 358).

[36] Sober (2015, 267). Others argue the same point (Shafer-Landau 2003, 110–14).

[37] Mogensen (2014, 108–12) presents a number of considerations for why they nonetheless shouldn't.

Ockham's razor, in the form of the anti-superfluity principle, therefore, does not, on its own, succeed in explaining how the genealogy of moral beliefs suffices to undermine them wholesale.

3.3.2 Explanatory Dispensability

There is a different way in which explanatory dispensability could explain how the evolutionary genealogy of moral beliefs could uniformly undermine them. This could be held to stem from the fact that the truth of the content of our moral beliefs fails to be implied by their best explanation. One could hold, like Harman, that this in and of itself is sufficient to undermine moral beliefs.

Joyce seems to rely on some such idea, even if inchoate in form. Formulated as a general principle, and applying it to belief-formation processes and ultimate causes, rather than individual beliefs and proximate causes, such a principle could look like this.[38]

> EXPLANATORY DISPENSABILITY
> For any belief-forming method M, if the explanation of why M has been selected for makes no indispensable reference to M-beliefs being true, then S' beliefs formed through M are undermined.[39]

This principle inherits Harman's requirement that moral beliefs (or here, moral belief-forming processes) must in some way be explained by an indispensable appeal to moral facts in order to be in epistemic good standing (e.g., be reliable).

The second answer to how evolutionary debunking arguments debunk is therefore that the belief-forming processes that generate moral beliefs have been selected for in a way that does not require the postulation of moral facts. This lack of explanatory indispensability is taken to be sufficient to establish that the process is epistemically defective and therefore sufficient to undermine moral beliefs.

[38] For discussion of debunking arguments targeting belief-forming processes rather than sets of beliefs, see Nichols (2014).

[39] Braddock (2017, 98) interprets Joyce to be concerned with process-insensitivity in this way. Cf. Mogensen (2014, 72). As discussed previously, this principle needs to be amended to also cover dispensability for proximate explanations. From here on out I will drop such qualifications.

One might worry that this principle is a bit too abstract or general to be informative, or at least that it lacks a proper defense. One would want to know *why* it is that such explanatory dispensability suffices to undermine moral beliefs. To this end, it would be necessary to determine exactly what kind of relation *is* needed between the relevant facts and our beliefs about them, and also why the lack of such a relation is sufficient for undermining the justificatory status of the output of such a belief-forming process.

One might think that one can avoid the need for answering such questions. After all, while epistemologists often disagree about *why*, exactly, a given belief-forming process is defective, they nonetheless agree that processes such as "confused reasoning, wishful thinking, reliance on emotional attachment, mere hunch or guesswork, and hasty generalization" *are* defective.[40] If moral beliefs could be shown to be produced by some such defective-by-consensus process, that could be sufficient to show *that* the process is epistemically defective without explaining *why*.[41]

I believe that this hope is overly optimistic in the case of the belief-forming processes underlying moral belief. One reason for this is the possibility of epistemological maneuvering. Alvin Plantinga has argued that even wishful thinking need not be an epistemically deficient belief-forming process given appropriately outré background assumptions.[42] We will also discuss accounts of moral beliefs that similarly explain why explanatory dispensability might not matter in later chapters.

If we instead attempt to explain what is epistemically defective about belief-generating processes such as those described above, one could adopt the principle discussed in the next subsection.

3.3.3 Modal Conditions

As should be clear by now, an important strain in Joyce's argument, as well as Harman's, is that there is something epistemically deficient about beliefs whose truth is not implied by their best explanation. When attempting to determine exactly what it is about this fact that renders them epistemically deficient, Joyce at times suggests that it is their insensitivity to evidence.

[40] Goldman (1979, 9).
[41] Nichols (2014, 733).
[42] Plantinga (2000, 195–98).

> On the assumption that my favored hypothesis about the "moral sense" is correct, it follows that the process by which humans form moral beliefs is an unreliable one, for they are disposed to do so regardless of the evidence to which they are exposed.[43]

Belief-forming processes that are insensitive to evidence tend to fail to reliably track the relevant truths, meaning that there will be many cases (either actual or counterfactual) where such a process produces false beliefs. For instance, if I form my beliefs about next week's weather by consulting tea leaves, this process will be insensitive to meteorological evidence and is therefore likely to (either actually or counterfactually) quite often get things wrong.

When made more precise, such a notion of epistemic sensitivity (to facts or evidence) is usually cashed out by appealing to a counterfactual condition drawn from the work of Robert Nozick.[44]

> SENSITIVITY
> S's belief that p, formed via method M, is sensitive if and only if, if p were false, S would not believe that p via M.[45]

The gist behind epistemic sensitivity is the requirement that an agent should be able to discern whether a belief is true or false. If we form beliefs in a way that fails to be sensitive, we would hold these beliefs independently of whether they were true. Since we would hold them even if they were false, having insensitive beliefs therefore tends to indicate that the beliefs are unreliable. This might be thought to be a clear epistemic defect of a belief. Some, such as Nozick, have argued along these lines that a belief must be sensitive to constitute knowledge, others have argued that a lack of sensitivity is sufficient to undermine justification.[46]

Joyce argues that since the belief-forming process underlying our moral beliefs is explained by facts about what is fitness-enhancing, moral beliefs will not be sensitive to evidence concerning what is morally right or wrong.

> Suppose that the actual world contains real categorical requirements [...] In such a world humans will be disposed to make moral judgments [...] for natural

[43] Joyce (2001, 162–63).

[44] Nozick (1981).

[45] I borrow the formulation of sensitivity from Bogardus (2016, 639).

[46] For sensitivity as a necessary condition for knowledge, see Nozick (1981, 179). As a sufficient condition for epistemic defeat, see Braddock (2017, 93–94), cf. Mogensen (2014, 112–13).

> selection will make it so. Now imagine instead that the actual world contains no such requirements at all—nothing to make moral discourse true. In such a world humans will still be disposed to make these judgments […] just as they did in the first world, for natural selection will make it so. What this shows is that the process that generates moral judgments exhibits an independence relation between judgment and truth, and these judgments are thus unjustified.[47]

Here is how the rough argumentative structure of a global moral debunking argument based on a modal condition such as insensitivity could go. If one takes knowledge to imply sensitivity, then showing that moral beliefs fail to satisfy SENSITIVITY would show that we do not possess moral knowledge. If lack of sensitivity is taken to be a sufficient condition for epistemic defeat, then showing that moral beliefs fail to satisfy SENSITIVITY would show that moral beliefs are uniformly undermined. We will return to evaluate such arguments from insensitivity, as well as related modal conditions, in Chapter 6.

The third answer to how evolutionary debunking arguments debunk is therefore that the undermining force of genealogical explanations of moral beliefs consists in claiming that moral beliefs fail to have a certain modal profile that is either required for knowledge or the lack of which is sufficient for epistemic defeat.

3.3.4 Explanatory Constraints

In later work, Joyce has acknowledged that his earlier argument contained an implicit reliance on epistemic sensitivity as spelled out in the previous subsection. He now puts less emphasis on the parts of his argument that relied on such modal conditions and is now more inclined towards adopting some form of explanatory constraint on justification, in the form of a sufficient condition for epistemic defeat. Joyce is again taking inspiration from Harman, in thinking that the lack of a mechanism, or an account, that makes sense of the explanatory relevance of moral facts can undermine a belief. Joyce quotes Harman approvingly as stating that

> what's needed is some account of *how* the actual wrongness of [something] could help explain [someone's] disapproval of it. And we have to be able to believe in this account. We cannot 'just make something up.'[48]

[47] Joyce (2001, 163). Others who have pressed a sensitivity based debunking argument include Ruse (1986, 254) and Sinnott-Armstrong (2006, 43).

[48] Harman (1986b, 63), quoted in Joyce (Joyce 2016b, 133). Emphasis and insertions in the original.

Such a demand is subtly different from the requirement that moral facts be explanatorily indispensable or that our beliefs about them be epistemically sensitive. This explanatory requirement is much weaker, in that all it requires is a plausible account of how there could be *some* explanatory connection between the moral facts and our beliefs about them. The constraint therefore requires a *believable account* of how moral facts can participate in the explanation of our beliefs or other attitudinal states. In other words, it would seem to pose something like the following explanatory constraint.

> EXPLANATORY CONSTRAINT
> If S' belief (or other attitudinal state) that p is not explained by the fact that p, or we lack an account of this explanatory relation, then S' belief that p is undermined.

While Joyce does not go very far to motivate the constraint he envisions, we can supply some further steps in that direction. There seems, quite simply, to be something strange—perhaps even epistemically impermissible—about beliefs that are formed in a way that completely fails to trace back to the facts they are about. These are cases where you believe that p, but the explanation of *why* you believe that p in no way traces back to the fact that p. This is particularly acute in the moral domain. There is something distinctively odd—and perhaps even irrational—about a person who would claim that "Genocide is wrong, but my belief that genocide is wrong is in no way responsive to, caused by, or otherwise guided by the fact that genocide is wrong."[49]

Building a debunking argument around EXPLANATORY CONSTRAINT would involve arguing that a non-naturalist has no mechanisms or accounts available for explaining how it could be that the moral fact that p could explain our belief that p. We will return repeatedly to evaluate debunking arguments that attempt to defend such a claim in Chapters 5, 6, and 7.

The fourth answer to how evolutionary debunking arguments debunk is that we lack an account of how moral facts can enter into explanations of moral beliefs, and that the lack of such an account is sufficient for epistemic undermining.

We have now seen four ways in which Joyce's claim that "the availability of an empirically confirmed genealogy of those beliefs that nowhere implies or presupposes their truth" can be supported by an appeal to general principles that could be thought to uniformly undermine moral beliefs. Having done so,

[49] Cf. Fitzpatrick (2015, 896).

it is worth considering the relation between these principles and the appeal to empirical speculations concerning the genealogy of moral beliefs.

Joyce himself plays up the dialectical and epistemic importance of the evolutionary genealogy of moral beliefs being *empirically supported*.[50]

> Overlooking this may encourage one to consider this argument as unimpressively analogous to standard challenges from the philosophical skeptic. […] But the view under discussion here does not come so cheap […] The argument does not depend on invoking extreme standards for epistemic justification; the skeptic is not requiring people to consider outlandish brain-in-vat-type possibilities that they would ordinarily scoff at. If the everyday standards for being morally justified take account of empirical data concerning human evolution, then if these data ultimately show moral beliefs to be unjustified it will be by ordinary epistemic standards.[51]

Evolutionary influences might therefore be a particularly perspicuous example of how humans have come to possess an epistemically deficient belief-forming process. It also rebuts the objection that a debunking argument is merely a “what if” skeptical scenario.

Despite this, the principles that would seem to undergird Joyce’s debunking argument are not dependent on the details of evolutionary theory. An exception might be EXPLANATORY DISPENSABILITY, but even in that case, the central issue is whether the lack of a certain type of explanatory relation is sufficient for epistemic defeat. At best, empirical findings from evolutionary theory might show that the belief-forming processes generating moral beliefs do lack this explanatory relation. But the principle is fully general.

Empirical details are even less obviously needed when the argument is directed at a non-naturalist, as they have been thought to lack any believable account or mechanism of how the belief-forming process could be explanatorily connected to the contents of the beliefs it outputs. If the lack of such an explanatory connection is sufficient for epistemic defeat, and the non-naturalist is unable to satisfy it, then the empirical details of selection processes fade into the background.

Joyce has come to much the same conclusion.

> [T]he evolutionary perspective is, strictly, dispensable. Were we to explain our moral beliefs by reference to, say, developmental and socialization processes, then, so long as these processes similarly nowhere imply or presuppose that our

[50] Joyce (2006, 187–88)

[51] Joyce (2006, 187–88).

> or anyone else's moral judgements are true, the same epistemological conclusion could be drawn.[52]

Lastly, it is worth noting that Joyce's considered version of his argument—couched in terms of the EXPLANATORY CONSTRAINT—is not restricted to targeting only non-naturalists, or even only moral realists. Joyce claims that his argument confronts any "moral success theorist," where that term covers anyone who thinks that we routinely have true, justified moral beliefs. This category therefore includes non-skeptical anti-realists, such as expressivists and constructivists, as well as realist views.[53]

In the next chapter, we will look at a different, and conditional, evolutionary debunking argument that expressly targets realist views, while holding their susceptibility to debunking arguments as a point in favor of non-skeptical, anti-realist views.

3.4 Summary & Preview

In this chapter, I have argued that Joyce's evolutionary debunking argument can be interpreted as employing a number of different epistemological principles in order to secure its conclusion. We have seen that, unlike Harman, Joyce is not necessarily committed to the claim that certain types of explanatory connections are required for generating defeasibly justified belief. Rather, Joyce seems to defend the view that the absence of such explanatory connections is a sufficient condition for epistemic defeat.

We saw that Joyce took Ockham's razor to show that non-natural moral facts are explanatorily superfluous and that our belief in them is thereby undermined. Similarly, we found that he can be understood as holding that belief-forming processes, the selection for which makes no indispensable reference to the truth of the content of the outputted beliefs, are epistemically defective.

Alternatively, he might be understood as employing a modal condition as either a necessary condition on knowledge or as a sufficient condition for epistemic defeat. Lastly, Joyce has more recently adopted a requirement from Harman that states that we need a believable account of how moral facts could explain our belief into an explanatory constraint. These principles—bracketing the modal condition for now—were as follows.

[52] Joyce (2016b, 125).
[53] Joyce (2016b, 126–27).

Explanatory Principles

> Ockham’s Razor (Anti-Superfluity)
> Avoid positing theoretical components which are not required for the purpose of explaining the relevant data.
>
> Explanatory Dispensability
> For any belief-forming method M, if the explanation of why M has been selected for makes no indispensable reference to M-beliefs being true, then S’ beliefs formed through M are undermined.
>
> Explanatory Constraint
> If S’ belief (or other attitudinal state) that p is not explained by the fact that p, or we lack an account of this explanatory relation, then S’ belief that p is undermined.

I concluded that Ockham’s razor was not obviously applicable to moral facts, given the conception of such facts endorsed by most non-naturalist (as well by many other realists). This was because the principle does not apply in the context of non-empirical, non-data-driven theorizing.

That leaves the two principles that focus on explanatory connections. The attempt to ground debunking arguments in the absence of certain—or even any—explanatory relations will be a running motif in the chapters to come. In Chapter 6, we will consider ways of developing and refining Explanatory Dispensability. The explanatory constraint Joyce picks up from Harman—Explanatory Constraint—will make an appearance in one form or another in both Chapters 5, 6, and 7. In Chapter 5, it will be discussed in connection with Street’s debunking argument, which will be introduced in the next chapter. In Chapter 6, we will look at a response to debunking arguments that attempts to deny the principle outright.

In Chapter 7, we will look at an attempt to formulate a refined explanatory constraint, as well as a rationale for why a belief’s explanatory connections take priority over its modal profile when it comes to determining reliability.

In addition to these explanatory principles, we also saw that Joyce can be reasonably interpreted as relying on a modal condition.

> SENSITIVITY
> S's belief that p, formed via method M, is sensitive if and only if, if p were false, S would not believe that p via M.

The worry that moral beliefs fail to satisfy certain modal conditions, such as SENSITIVITY, and therefore either fail to constitute knowledge or have their justification defeated is explored in Chapter 6. There, we look at attempts to capture the problematic aspect of beliefs that are insensitive or otherwise have an epistemically unhealthy modal profile.

4 Street's Darwinian Dilemma

4.1 Introduction

This chapter considers a second global evolutionary debunking argument, due to Sharon Street. Unlike Joyce, Street does not seek to ultimately undermine the justificatory status of our moral beliefs. Instead, she seeks to show only that *given a particular conception* of moral facts—moral realism—moral beliefs are undermined. In fact, Street makes this claim for the wider class of *normative beliefs* and claims that they are undermined conditional on adopting what she calls *normative realism*. The structure of Streets debunking argument makes it a conditional debunking argument, as I described them in §1.3.

I start by setting out the background for Street's dilemma, which concerns the extent to which evolutionary pressures have shaped our normative attitudes (§4.2). I then consider the two horns of the dilemma. The first horn holds that the evolutionary influences on normative beliefs have guided us in the direction of having true normative beliefs (§4.2.1). Street argues that on this horn, a realist account—and especially a non-naturalist realist account—will always be defeated by other, non-truth tracking accounts of the relation between evolutionary influences and our beliefs. Street holds that the latter type of accounts will win out on the basis of empirical considerations and ordinary factors at play in theory choice.

The second horn denies that evolutionary influence has guided us in the direction of having true beliefs (§4.2.2). Street argues that this leaves the realist embracing an implausible coincidence that begs for an explanation. This coincidence is the degree of overlap between the content of our normative beliefs and the content of the purported normative truths. If evolutionary influences have significantly shaped our normative beliefs, but not guided us in the direction of having true such beliefs, how could we claim to have a sizable amount of true normative beliefs?

Having set out the two horns, I then consider suggestions for how to understand Street's worry about the epistemic coincidence normative realists face on the second horn of her dilemma (§4.3). In particular, how this coincidence could be taken to secure the conclusion that normative beliefs, on a realist conception, would be uniformly undermined. I end by providing a summary of the chapter and restating the general principles I have taken to support Street's debunking argument (§4.4).

4.2 The Dilemma

In a much-discussed paper, Street argues that the evolutionary influences on our normative judgments raise a 'Darwinian Dilemma' for *normative realists* about practical reasons.[1] On Street's conception, normative realism is characterized primarily by the commitment that "there are at least some evaluative facts or truths that hold independently of all our evaluative attitudes."[2] Street uses 'evaluative fact' and 'evaluative truth' as catch-all terms for practical normative facts, including moral facts, while she intends 'evaluative attitudes' to capture both cognitive and conative states, including both conscious and unconscious normative judgments as well as desires, and attitudes of approval and disapproval.[3]

Street's argument casts a wider net than Harman's and Joyce', as it targets a broader swath of normative discourse than merely the moral domain. Street takes it to target realist views of all of practical normativity, including the prudential domain, and has elsewhere expanded it to target epistemic normativity as well.[4] In later work, she argues that her argument even targets non-realist views, such as normative quasi-realism.[5] In this chapter, we will mainly be concerned with the impact of Street's argument for the moral domain, and as targeting a non-naturalist, but in later chapters, we will return to the broader reach of her argument.

Street's evolutionary debunking argument, like most others, begins with the claim that evolutionary selection pressures have significantly influenced our normative beliefs. If true, this saddles the realists with the explanatory obligation to answer the following question: What is the relation between the evolutionary influences on the content of our normative beliefs and the purported stance-independent normative facts?

Street sets out two ways in which the realist could discharge this explanatory debt. First, they could claim that evolutionary influences somehow *track*

[1] Street (2006).
[2] Street (2006, 110).
[3] Street (2006, 110).
[4] Street (2009).
[5] Street (2011).

the normative truths, and therefore "guide" us in the direction of having true normative beliefs. Alternatively, they could claim that such evolutionary influences bear *no relation* to normative truths, and that to the extent that our normative judgments are true, it is no thanks to such influences. Street claims that whichever option the normative realist opts for, the resulting realist view is ultimately untenable.

In another departure from Joyce's argument, Street does not take the conclusion of her debunking argument to be that our normative (and therefore moral) beliefs are unjustified *tout court*. In the taxonomy set out in §1.3, Street is defending a conditional debunking argument, as she thinks her dilemma rules out certain metanormative views, such as normative (and therefore moral) realism.[6] Even so, she holds that certain non-skeptical metanormative *anti-realist* accounts can successfully answer the Darwinian dilemma. For instance, she claims that *normative constructivism* could successfully account for the evolutionary influences on normative beliefs and therefore avoid skepticism.[7] This feature of Street's argument will become especially important when we discuss the threat of self-defeat facing debunking arguments in Chapters 7 and 8.

The conclusion of Street's argument is therefore not that we should think, like Joyce holds, that all moral beliefs are unjustified. Rather, it is that since normative realism fails, we should opt for a different, and better, metanormative theory. We will return to this issue after considering her dilemma.

I will now go through Street's empirical premise and the two horns of her dilemma in turn. As for the empirical side of things, Street presents a speculative sketch of the mechanisms through which evolutionary influence on normative cognition could take place. Like Joyce, Street is aware of the implausibility of claiming that the content of normative judgments is itself capable of being heritable, and so it is implausible that there could be systematic selection for such content.[8] She instead proposes that what might be heritable is instead 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 favor of' something else."[9]

[6] By 'metanormative view' I mean a position that concerns the multiple subdomains within, or the whole of, the normative domain, including morality, as opposed to a metathetical account of merely the moral domain.
[7] Street (2006, 152–54). Street is not alone in thinking views like constructivism can more easily avoid epistemological challenges such as evolutionary debunking arguments (cf. Enoch 2009, 324; Finlay 2007, 829).
[8] Street (2006, 118–19).
[9] Street (2006, 119).

Street puts forth a sketch of how such a trait could be selected for. Certain forms of behavior tend to be fitness-enhancing. This includes taking care of one's children and avoiding serious harm. However, hardwiring behavior or propositional content is a relatively difficult and inelegant way of fostering such behaviors. Street therefore focuses on the fact that having the motivational tendencies that lead to performing such behaviors is fitness-enhancing as well. Having these motivational tendencies, and in turn, engaging in these behaviors, made it more likely that ancestors would survive and reproduce.

As motivational dispositions are a better candidate for being a heritable trait than either hardwired behavior or propositional content, they gradually became more prevalent in the population and so did the behaviors they promoted. Such motivational tendencies, on Street's sketch, could have resulted in a form of proto-normative judgments, such as feeling compelled to punish cheaters or to engage in cooperative behavior. Over time such sub-conscious dispositions evolved into the capacity for full-blooded, linguistically expressed normative judgments.

On this account, natural selection did not influence the content of our normative judgments directly, but rather through shaping our motivational and behavioral tendencies in a way that inclined us towards making normative judgments with a particular type of content.[10] Such a story is further supported, Street claims, by the observation that our normative judgments do conform, in rough outline, to what we would expect them to be if they *were* significantly shaped by selective forces.

This means that normative realists, who subscribe to a domain of stance-independent normative facts, face a choice between three options. First, they could deny the empirical claim that the content of our normative beliefs has been subject to significant evolutionary influence, either directly or indirectly.[11] This might seem implausible on the face of it, however. If the realist instead accepts that there has been such influence, they are caught by Street's dilemma and must either *assert* or *deny* that there is some relation between the evolutionary influences on our normative judgments, on the one hand, and the stance-independent normative truths on the other. Either way, the argument goes, the resulting realist view is untenable. I will now go through these two horns of the dilemma.

[10] Street (2006, 158–59 n. 22).

[11] For discussion or defense of this possibility, see Nagel (2012, 105), Huemer (2016), Parfit (2011, sec. 119), FitzPatrick (2015), and Isserow (2019).

4.2.1 First Horn: Asserting a Relation

The type of relation a realist would need to commit to is one that explains how the influence of evolutionary processes has led to humans possessing true normative beliefs. By asserting that there is some such relation, the realist is then tasked with elucidating what this relation consists in. To see what this relation is meant to explain, and what it could look like, it might be worth considering an analogy not employed by Street.

A realist about the external world arguably faces the same 'Darwinian challenge' as the normative realist.[12] No one should deny that our perceptual faculties have been strongly shaped by evolutionary pressures.[13] We also take ourselves to have a significant number of true beliefs formed on the basis of perception. Since we think there are stance-independent facts about the world we perceive around us, we need to explain how it is that evolutionary pressures have led to us having a visual system that allows us to form true beliefs about the world around us.

To see what such a relation could look like, the external world realist could point to the fact that beliefs about physical objects tend to be formed in response to causal stimuli. Perceptual beliefs tend to causally track facts about our surroundings. Our perceptual system has been selected for on the basis of its ability to track facts about mid-size objects at a moderate distance. The truth or falsity of perceptual beliefs is therefore not irrelevant to their ability to increase the perceiver's relative fitness. As a result, there is a plausible story to be told about why the evolutionary pressures on the human perceptual system would dispose us to form beliefs the content of which overlap with the independent facts about physical objects.

It is of course *possible* to create skeptical challenges to this view as well, but the point here is that the realist simply needs an account of the relation that beats out other accounts. That seems true of the external world realist's account.

The proposed causal relation serves to explain two things. First, it explains how there could be a convergence between the content of perceptual beliefs as shaped by evolution and the stance-independent facts about physical objects. Second, it also makes it plausible that such convergence has occurred *because* our perceptual beliefs are true. That is to say, the selection for the

[12] We discussed a similar issue for realists about the objects of perception in §1.2.
[13] Stevens (2013).

perceptual beliefs (or rather, the perceptual system outputting these beliefs) has taken place *because* the resulting beliefs are sufficiently veridical.

Street considers the prospects for proposing an analogous account of the relation between normative beliefs and normative facts. Street calls it the *tracking account*. She imagines the normative realist elaborating on such an account by claiming that

> 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.[14]

The tracking account holds that our normative beliefs are to be explained by the normative facts themselves, much like how the facts about physical objects explain our beliefs about them. We have the normative beliefs that we do, on this account, because holding true normative beliefs is evolutionarily beneficial.

The tracking account would allow the normative realist to explain two things. First, it would explain how there could be a convergence between the content of normative beliefs and the stance-independent normative facts. Second, it also makes it plausible that such convergence has occurred *because* the normative beliefs are true. That is to say, the selection for normative beliefs (or rather, for our normative motivational tendencies) has taken place *because* the beliefs in question are sufficiently veridical. This, in turn, is explained by the fact that having veridical normative beliefs is evolutionarily beneficial.

The resulting tracking account therefore holds that

> the presence of [normative] judgements is explained by the fact that these judgements are true, and that the capacity to discern such truths proved advantageous for the purposes of survival and reproduction.[15]

As Street sets it out, the tracking account attempts to explain not only why the evolutionary processes have led us to have true normative beliefs, but also why it is *the truth of the beliefs* that has conferred the evolutionary benefit. These two elements—being disposed to have true beliefs and having the beliefs confer a relative fitness advantage in virtue of being true—can come apart. Street nonetheless considers them as package deal.[16] Before considering

[14] Street (2006, 126), cf. Nozick (1981, 337).

[15] Street (2006, 126).

[16] As Mogensen (2014, 72–73) points out, Street shifts between describing the tracking account as one that merely attempts to explain why selective pressures has made us disposed to have true moral beliefs, and of it also claiming that it is the beliefs being true that explain their being

the possibility of these elements coming apart, let us set out Street's dilemma as she conceives it.

The tracking account presents a hypothesis that is in the business of explaining empirical data. It attempts to provide an ultimate cause for patterns of normative judgments found in the human population. It does so by holding that it confers a relative fitness advantage to be able to grasp normative truths.

Street claims that any such version of the tracking account will be empirically implausible and explanatorily inferior to an account of our normative beliefs which jettisons the commitment to a domain of stance-independent normative facts.

To see the force of the first horn of Street's dilemma, consider how it would apply to a normative non-naturalist. Such a non-naturalist would be required to explain why the beliefs that the realist holds to be true would have been fitness-enhancing for our ancestors, where the fitness advantage is supposed to be explained in terms of the truth of the beliefs' content.

This is again parallel to how it is the truth of our perceptual beliefs that explain why they confer a fitness-advantage. Parfit, who—as explained in §1.4—is *almost* a non-naturalist on my characterization, attempts a sketch of what such a view could look like, when applied to reasons for belief.

> When some fact gives us a reason to have some belief, this normative property of being reason-giving is not an empirically discoverable feature of the natural world. Nor could we be causally affected by such normative properties. But natural selection might explain how, without any such causal contact, our ancestors became able to respond to such reasons, because that enabled them to form many true beliefs about the world, some of which helped them to survive and reproduce. If natural selection explains how early humans acquired this ability to respond to reasons, it would not be a coincidence that these humans could form these many true beliefs.[17]

The emerging picture is therefore as follows.

> Just as the faster cheetahs and taller giraffes tended to survive longer and have more offspring, who inherited similar qualities, so did the humans who were better at reasoning validly and responding to reasons.[18]

selected for. It seems reasonable to read her along the lines of the latter, more demanding, version of the account as that is the view her arguments target (cf. Street 2008b, 210).

[17] Parfit (2011, 496–97). As is clear from the quote, Parfit is here discussing normative properties, facts, and reasons more generally, including not only moral but also epistemic reasons.

[18] Parfit (2011, 494).

This would mean, for the normative non-naturalist, that having beliefs about causally inert, non-natural normative facts systematically conferred a selective advantage *in virtue of the truth of their content*. This is a radical suggestion. Textbooks on evolutionary theory would need to be revised in order to include the selective advantages of this type of cognitive mechanism. There is presently no suggestion as to how such a cognitive mechanism could work, and many reasons to take issue with its plausibility.[19]

As an empirical theory, Street believes, the tracking account is untenable. This is because there will be a more scientifically respectable theory available that explains the distribution of normative beliefs in the human population without needing to explain how we are able to track non-causal and non-natural normative facts. Instead, such an alternative theory would consider the capacity and tendency for normative thought to be a biological adaption that arose because the basic evaluative attitudes (that in turn gave rise to full-fledged normative judgments) were selectively advantageous.

On such an alternative account, normative beliefs were not selectively advantageous *because* they are true, but rather because the mere possession of such beliefs promotes survival and fosters cooperation.[20] Street claims that this latter view, which she terms the *adaptive link account*, has a number of significant advantages over the tracking account.

First, the adaptive link is more parsimonious in that it explains all that the tracking account does, without postulating stance-independent normative truths, the knowing of which confers a relative fitness advantage. Second, the adaptive link account is clearer, as it only appeals to mechanisms and relations that are already in place for explaining other types of cognitive faculties and abilities. The answer to why we make the normative judgments we do, on this account, is in large part that "ancestors who judged that they should care for their offspring met with greater reproductive success simply because they tended to care for their offspring—and so left more of them."[21]

Third, the adaptive link account does a better job of explaining what is at issue—the pattern of our observed normative judgments. Street thinks the adaptive link account—unlike the tracking account—can provide a straightforward answer to a number of questions of interest on this point. For instance,

[19] One candidate mechanism would be the possession of quasi-perceptual rational intuition, but as discussed in §3.3.1, such an account is highly unattractive. Furthermore, such an account would not suffice to explain why it is the truth of a belief's content, rather than the mere possession of the belief that confers a fitness advantage.
[20] Street (2006, 127–28). This is a simplification of the account.
[21] Street (2006, 131–32).

it provides a better answer to the question of why so many of the normative judgments we make are exactly those that we would make if our judgments had been selected for.[22]

The adaptive link account, Street concludes, is in the end the more parsimonious, clearer, and explanatorily richer theory of why certain normative beliefs became widespread in the human population over time.

While realists other than non-naturalists could perhaps attempt to engage Street's arguments on these points, the non-naturalist is in a much less favorable position. Non-naturalists therefore tend to deny that it is the truth of our normative beliefs that explain why they are evolutionarily advantageous. When understood as set out above, the non-naturalist should therefore reject the tracking account, and opt for the second horn of the dilemma.

4.2.2 Second Horn: Denying a Relation

Faced with the difficulties involved in defending an empirical hypothesis against low-cost alternatives, a non-naturalist might instead opt for the second horn of Street's dilemma, by denying that holding true normative beliefs is evolutionary beneficial in virtue of their truth. This means that the realist is thereby freed from the task of explaining the relation between the influences on our moral beliefs and the stance-independent normative truths.

By claiming that there is no such relation, however, one would need to explain how we have nonetheless come to possess true normative beliefs. After all, the non-naturalist has now granted that normative beliefs *are* significantly influenced by evolutionary forces, and also that such forces are not influencing us in the direction of having the beliefs we do *because they are true*. On this horn of the dilemma, the truth or falsity of normative beliefs is taken to be immaterial to the process of selection that has disposed us to have them. A substantial overlap between the content of normative judgments produced by such a process and the now explanatorily isolated normative facts might seem, as Street claims, too much of a coincidence to be believed.

There are at least two possible interpretations of what it is, exactly, that Street finds coincidental. On the first interpretation, which has been suggested by Mogensen to be the central one, the problematic coincidence is "that beliefs that overlap with the objective normative facts have proven reproductively

[22] Street (2006, 132–33).

advantageous."[23] If the tracking view were correct, this would be no surprise. But as the non-naturalist has given up such an account, it now becomes unclear how it came to be that the true normative beliefs, as it happens, also happen to be selectively advantageous. Mogensen has argued that this coincidence need not be problematic.[24]

There is also another, more straightforward possibility, which is that the coincidence has to do with explaining the overlap of the content of our normative beliefs with the content of stance-independent normative facts.[25] This option is discarded by Mogensen, because he believes it falls prey to his criticism of debunkers' failure to distinguish proximate and ultimate levels of explanation.[26] His reasoning for this is that if the realist can show that normative beliefs have normative facts as part of their *proximate cause*, then Street's argument would have no force when interpreted in this way.

Non-naturalists—because of their embrace of non-natural, non-causal normative facts—are not well positioned to employ normative facts as a proximate cause for their beliefs. For this reason, *pace* Mogensen, the second interpretation of Street's argument is in fact perfectly employable, at least when targeting a non-naturalist. Such theorists do face a significant obstacle when attempting to explain the mere convergence of the content of our normative beliefs with the stance-independent normative facts.

How can the non-naturalist respond at this juncture? One might worry that there is not much that can be said other than to chalk it up to good fortune that the content of the beliefs, which evolution has disposed us towards believing for entirely orthogonal reasons, and the stance-independent normative facts overlap.[27] This might, as Street suggests, seem too much of a coincidence to be believed. Are we really to accept that the evolutionary influences on our beliefs have caused certain beliefs to become widespread for reasons having nothing to do with their truth, and that those beliefs are, as it happens, true?[28]

> [H]ow incredible (not to mention how extraordinarily convenient for the realist) that, as a matter of sheer coincidence, a capacity happened to arise (as the entirely accidental byproduct of some totally unrelated capacity […]) which operates to grasp precisely the sort of independent truths postulated by the realist.[29]

[23] Mogensen (2015, 201).
[24]Mogensen (2015, 201).
[25] Cf. Street (2008b, 211).
[26] We discussed this criticism §2.3.
[27] Dworkin (1996, 125–27) seems to argue for something like this.
[28] Cf. Street (2006, 132).
[29] Street (2006, 142–43).

Street thinks that postulating any such coincidence is sufficient to uniformly undermine normative belief—or even render it false—given a realist construal. Much like Joyce, Street opts for an evocative analogy to bring out the undermining potential of her argument.

> [A]llowing our evaluative judgements to be shaped by evolutionary influences is analogous to setting out for Bermuda and letting the course of your boat be determined by the wind and tides: just as the push of the wind and tides on your boat has nothing to do with where you want to go, so the historical push of natural selection on the content of our evaluative judgements has nothing to do with evaluative truth.[30]

Street's worry is that it is almost inconceivable that we have, through dumb luck, formed true beliefs *and* that these beliefs are just the beliefs we would have formed (and in fact have formed) on the basis of evolutionary processes that "blindly" promote evolutionary fitness. Because of this, Street claims that opting for the second horn of the dilemma undermines our normative beliefs.

Street sometimes even seems to hold that it supports the stronger thesis that the beliefs in question are *likely false*.[31] Relatedly, she at times formulates the second horn of her dilemma in terms of the evolutionary influences on normative beliefs constituting a *distorting* influence, and thereby claiming that, on a realist view, it is *unlikely* that we have true beliefs, or conversely, that we are likely to have false beliefs.[32] For convenience, I set such issues aside and prioritize her claim that "I'm understanding 'denying a relation' to be a matter of regarding the influence of evolutionary forces on our evaluative attitudes as no better than random with respect to the truth."[33]

Having set out Street's worry about the coincidence needed to explain the overlap between the content of the normative truths and of the beliefs evolutionary influences have disposed us towards, let us now look at *how* such a coincidence could threaten to uniformly undermine moral beliefs.

[30] Street (2006, 121).
[31] E.g. Street (2006, 125).
[32] Street (2006, 121).
[33] Street (2008b, 208).

4.3 Epistemic Coincidences: Modal Variability and Unexplained Reliability

Focusing now on moral beliefs, a moral non-naturalist, on the second horn of Street's dilemma, is faced with the following task. She must explain how the content of the stance-independent moral truths has come to overlap with the content of our moral beliefs, given that the latter was significantly influenced by processes that are explanatorily disconnected from the moral truths.

Could the non-naturalist merely chalk it up to a lucky coincidence? When does a coincidence become *too much to be believed*? There is, after all, no fully general ban on believing in coincidences. What we would want is some principled criterion that spells out which coincidences are epistemically problematic, in the sense that they are sufficient for removing a belief's positive epistemic status. Street is not very explicit about where to draw this line, only that a realist's embrace of the coincidence would cross it.

> This degree of overlap between the content of evaluative truth and the content of the judgements that natural selection pushed us in the direction of making begs for an explanation. Since it is implausible to think that this overlap is a matter of sheer chance—in other words, that natural selection just happened to push us toward true evaluative judgements rather than false ones—the only conclusion left is that there is indeed some relation between evaluative truths and selective pressures.[34]

What makes a coincidence 'beg for explanation' or be *too* implausible? Such questions, as well as how to interpret Street's worry about coincidences, have given rise to much debate.[35] Here, I will consider two ways of making Street's worry about coincidences more precise.

In one attempt to spell out what is problematic about the non-naturalist's reply, Street points to the vast realm of logically possible ways normative reality could have manifested.

> Of course it's *possible* that as a matter of sheer chance, some large portion of our evaluative judgements ended up true, due to a happy coincidence between the realist's independent evaluative truths and the evaluative directions in which natural selection tended to push us, but this would require a fluke of luck that's not only extremely unlikely, in view of the huge universe of logically

[34] Street (2006, 125).

[35] Bedke (2009; 2014); Kahane (2011); Mogensen (2015); Talbott (2015); Hopster (2019); Baras (2020).

> possible evaluative judgements and truths, but also astoundingly convenient to the realist.[36]

Street here attempts to bring out all the ways in which normative reality *could*, as a matter of logical possibility, have turned out. On such a modal picture "survival might be bad, our children's lives might be worthless, and the fact that someone has helped us might be a reason to hurt that person in return."[37] Given the vastness of such possibilities, it might seem incomprehensible that a truth-indifferent biological process shaped us to have a significant proportion of just those moral beliefs whose content happens to be true.

While Street does not suggest any particular principle for how to turn modal normative variation into a debunking argument, we have already considered one way of doing so. If Street is correct about the vastness of normative possibility, normative beliefs would likely fall prey to the epistemic sensitivity condition set out in the previous chapter. In Chapter 6, we will also look at a slightly different argument from modal variation as well.

Even disregarding the possibility of such modal variability, there are other ways to understand what is problematic with the non-naturalist's position. A popular suggestion has been that the coincidence noted by Street blocks the possibility of explaining the reliability of moral beliefs. In a later paper, Street recaps her argument by stating that the beginnings of a solution to her dilemma requires holding that

> evolutionary forces caused our moral beliefs to track the moral facts to an epistemically sufficient degree, and the realist is welcome to insist upon this. But the point of the Darwinian Dilemma is that the realist owes us some explanation of this fact. Is our ability to track the moral facts just a fluke? (I have argued this is implausible.) Does the tracking account provide an adequate explanation? (I have argued that it does not.) Is there some other good explanation? (I have argued that there is not, so long as one remains a realist about normativity.)[38]

This might make it seem that what we want, is an explanation of how it is that we have reliable moral beliefs. On a *very* simple construal, the reliability of a belief is explained only if there is some account of how it is that it corresponds

[36] Street (2006, 122). Italics in the original.
[37] Street (2008b, 208).
[38] Street (2008b, 2011).

to the truth.[39] David Enoch, taking inspiration from the Benacerraf-Field challenge against mathematical Platonism, has suggested that if there is no explanation of the reliability of moral beliefs, then that, in and of itself, is sufficient for undermining those beliefs.

This interpretation of Street is one where the epistemically malignant coincidence involved in the second horn of Street's dilemma is exactly the non-naturalists *lack of any explanation* for the epistemic reliability of moral beliefs. We can set this out more generally as follows.

> UNEXPLAINED RELIABILITY
> If the reliability of S' belief that p cannot be explained, S' belief that p is undermined.[40]

While temptingly simple, there are a number of issues to be sorted out concerning a principle such as this, including the notions of 'reliability' and 'explanation' employed. We will return to that task in the next chapter, when we consider a debunking argument built around UNEXPLAINED RELIABILITY.

Sometimes, of course, the explanation for why something is the case is that it is a complete fluke. A faithful interpretation of Street's argument would want to rule out that a non-naturalist can appeal to such "sheer chance" as an explanation for the reliability of our beliefs. We can therefore add the following principle.

> NO COINCIDENCE
> For any theory, T, which concerns a domain, D, if T entails that the reliability of D-beliefs is merely coincidental, that strongly counts against T.[41]

In the next chapter, we will consider whether UNEXPLAINED RELIABILITY and NO COINCIDENCE can be combined to form a successful debunking argument.

Before moving on, let us take note of a few aspects of Street's argument. First is the degree to which Street's evolutionary debunking argument deserves the name. She is explicit that the fundamentals of her argument do not

[39] Reliability is standardly attributed to belief-forming process, but the literature on debunking have tended to employ the term in different, but not altogether idiosyncratic ways. The relevant sense(s) of epistemic reliability is discussed in §5.3.1.

[40] This is a simplified version of the relevant principle. We will discuss various complications for it in the next chapter. For related principles, see Lutz (2020, 294) and Baras (2017a).

[41] For similar principles, see Bedke (2009; 2014) and Baras (2017a).

really depend on the particulars of evolutionary theory. As Street herself admits: "At the end of the day, then, the dilemma at hand is not distinctly Darwinian, but much larger."[42] As we have seen, this has been a motif for both the evolutionary debunking arguments we have looked at so far. The "evolutionary" aspect of evolutionary debunking arguments is therefore accepted to be inessential, although enticing and glossy, by most parties to the debate. Despite this, we will return to a form of debunking argument where the details of evolutionary theory are in fact essential in §6.4.

A second point concerns the dialectical structure of Street's argument. Street originally targeted only realist views and claimed that in view of the skeptical threat for realism, we should opt for some anti-realist view like normative constructivism. Some realists have responded that if Street's dilemma is sound, then embracing *realist skepticism* would be the appropriate response.[43] A possible upshot of Streets dilemma is therefore that one might subscribe to the existence of moral facts, while denying that we can have justified beliefs about them.

Street's conditional debunking argument is dialectically complex in ways that we will return to in later chapters. We will see that this provides the view with resources to answer important objections that face non-conditional debunking arguments.

4.4 Summary & Preview

In this chapter, we have looked at a second evolutionary debunking argument, which is conditional in nature. Like Joyce's argument, there is an attractive intuition behind the core of Street's argument. It is highly intuitive that postulating a massive coincidence to explain how we have come to have true moral beliefs would seem to undermine them.

We then looked at two ways in which Street's argument can be made more precise. On the one hand, Street makes a point of arguing that the logical possibilities for what the moral facts could have been made it unlikely that we would end up with true moral beliefs. This could allow Street to appeal to some modal condition—e.g., SENSITIVITY—to show that moral beliefs fail to satisfy either a necessary condition for knowledge, or a sufficient condition

[42] Street (2006, 155).

[43] Skarsaune (2011); Enoch (2011, 171–72).

for epistemic defeat. I will discuss the possibility that there are logically endless normative possibilities in Chapter 6.

We also saw that the worry about coincidence can be understood as the claim that the non-naturalist lacks an explanation of the reliability of moral beliefs. We saw that this can be captured by the combination of the following principles.

Reliability Principles

> UNEXPLAINED RELIABILITY
> If the reliability of S' belief that p cannot be explained, S' belief that p is undermined.

> NO COINCIDENCE
> For any theory, T, which concerns a domain, D, if T entails that the reliability of D-beliefs is merely coincidental, that strongly counts against T.

Beginning in Chapter 5, I will investigate a debunking argument constructed around UNEXPLAINED RELIABILITY and NO COINCIDENCE. This so-called reliability challenge will expand on Street's claim that there is something objectionably coincidental about moral beliefs on a non-naturalist conception. We will also consider a non-naturalist reply to such a challenge, which attempts to answer the challenge by invoking a third-factor explanation.

Chapter 6 considers a different response from the non-naturalist to the reliability challenge, which consists in showing that when NO COINCIDENCE is restricted to concern only *epistemically problematic* coincidences, it is not clear that it targets the non-naturalist who posits a cosmic coincidence to explain the reliability of moral beliefs.

We have now seen three paradigmatic debunking arguments set out in detail. A recurring motif in these arguments has been that they have lacked explicitly formulated principles that explain how genealogical information can uniformly undermine moral beliefs. Instead, they rely on analogical reasoning in order to support their epistemological conclusions.

Each of the three arguments we have looked at can be interpreted as relying on various principles which could serve as the theoretical backbone of a global moral debunking argument. Different debunking arguments, targeting different features of moral beliefs, can be constructed on the basis of these principles.

In Part II, I will evaluate the principles now set out, as well as more refined versions of them. I will argue for two important conclusions. The first is that no extant, non-conditional debunking argument succeeds. Debunking arguments, I show, have not seen the success many have predicted, because they are forced to take on far more epistemological and metaepistemological commitments than has hitherto been recognized. In the absence of fulfilling these tasks, debunking arguments are threatened by implausibility or self-defeat. I highlight important lessons that will help future debunking arguments mitigate this threat.

The second important conclusion is that there is a far deeper divide between the prospects for conditional and non-conditional debunking arguments than has hitherto been recognized. Conditional debunking arguments, like Street's, are threatened by entirely different challenges than non-conditional arguments. Conditional debunking arguments should defend a unified metanormative account that encompasses both the epistemic and the moral domain. Defending a debunking argument has proved difficult enough, without having to solve issues of metaepistemology on top. The general conclusion is therefore that the theoretical costs and obstacles for construing a successful debunking argument are far greater than has previously been recognized.

One could be forgiven for thinking that these conclusions merely indicate an uncharitable selection of principles, and that there must surely be better ones. That is not, I think, the case. I provide some considerations in support of this in the concluding chapter, where I set out conditions of adequacy that any successful debunking argument would need to satisfy. Satisfying them, I argue, will pose a significant challenge.

Part II

Evaluating Debunking Arguments

5 The Reliability Challenge: Why Third-Factor Explanations Fail

5.1 Introduction

In Part I, we explored three paradigmatic debunking arguments that sought to uniformly undermine moral beliefs, either conditionally or unconditionally, by an appeal to genealogical information. The previous three chapters also set out a number of candidate principles that could serve to explain how such an appeal to genealogical information could secure the intended conclusions of those arguments. I will now—in Part II—evaluate whether debunking arguments built on these principles, or refined versions of them, succeed.

In this chapter and the next, I evaluate a debunking argument centered on UNEXPLAINED RELIABILITY. Call this type of debunking argument the *reliability challenge*. In this chapter, I set out the reliability challenge and consider one popular strategy for answering it—the third-factor strategy. I argue that this strategy cannot succeed because of how non-naturalists should understand the structure of moral explanation. In the next chapter, I then consider a different strategy for answering the reliability challenge, which I argue does not fall prey to any such principled objections.

I begin this chapter by introducing the reliability challenge (§5.2) and elaborating on two notions central to it—reliability and correlations (§5.3). I describe an important correlation that any non-skeptical metaethical view must explain, and provide a taxonomy of possible models for explaining it.

I then set out a popular strategy for answering the reliability challenge—and explaining the correlation—namely the third-factor strategy (§5.4). I show that the literature surrounding third-factor explanations has been mired in confusion, which has resulted in a failure to properly distinguish genuine third-factor explanations from related, but importantly different, forms of explanations.

After having set out the structure of genuine third-factor explanations, I move on to consider one common objection to this strategy—that it is question-begging. I argue that, while true, it does not appear to be *problematically* question-begging and that pushing the issue will likely result in a legitimate dialectical stalemate (§5.5). I then go on to present a novel, principled objection to the third-factor strategy (§5.6). I argue that because of the metaphysical commitments of moral non-naturalism, the third-factor strategy cannot, even in principle, succeed. This leaves the reliability challenge unanswered.

5.2 Debunking as Challenging Reliability

In this chapter, we will continue to explore how the worry about coincidence on the second horn of Sharon Street's Darwinian Dilemma can be developed in a more explicit fashion. The beginnings of this task have been carried out by David Enoch, who has provided a highly influential interpretation of Street's argument. Enoch understands Street's worry about the coincidental nature of moral belief to be an instance of a challenge originally directed at mathematical Platonism, first by Paul Benacerraf and later in a refined version by Hartry Field.[1] It's worth setting out that challenge, before moving on to consider how it applies to the moral domain.

For a mathematical Platonist, the mathematical domain consists of abstract, irreducible, causally inert entities existing outside of space-time, as well as facts that are stance-independent, causally inert, and not reducible to non-mathematical facts. The so-called Benacerraf-Field Challenge highlights the epistemological difficulties such a view encounters when attempting to explain our epistemic success in engaging with the mathematical domain.

Here is Field's reconstruction of Benacerraf's original challenge.

> [I]f there are mathematical entities of the sort the platonist believes in (mind- and language independent, having no spatio-temporal location, unable to enter into physical interactions with us or anything we observe) then there seems to be a difficulty in seeing how we could ever know that they exist, or know anything about them; the platonist needs to explain how such knowledge is possible, and no answer is evident except one that posits mysterious powers of access to the platonic realm.[2]

The challenge, as formulated by Benacerraf, is for the mathematical Platonist to explain how it could be that we have acquired mathematical knowledge

[1] For Field's refined version of Benacerraf's (1973) original challenge, see Field (1989, 25–30). Schechter (2010) claims the reliability challenge to also be inspired by J.L. Mackie's argument from queerness (Mackie 1977).

[2] Field (1989, 25).

given that mathematical Platonism is true. Field goes on to modify the challenge such that it no longer concerns the *possibility* of having mathematical knowledge, or the *possibility* of having defeasible justification for mathematical beliefs. Instead, Field suggests that it should be understood as the challenge of explaining why it is that, when mathematicians believe some mathematical claim, that claim is very often true (and that when they disbelieve some mathematical claim, that claim is very often false).[3]

To see how it is possible to challenge the reliability of our mathematical beliefs without directly challenging the mere possibility of mathematical knowledge or defeasible justification for mathematical beliefs, it is worth quoting Field's own outline of the challenge.

> We start out by assuming the existence of mathematical entities that obey the standard mathematical theories; we grant also that there may be positive reasons for believing in those entities. These positive reasons might involve only initial plausibility [or that] the postulation of these entities appears to be indispensable.[4]

Field's version of the challenge begins by granting the mathematical Platonist the defeasible assumption that mathematical entities exist, as conceived by the Platonist, as well as defeasible justification for their mathematical beliefs. Having now granted the Platonist a fair amount outright, Field then presses the epistemological challenge.

> [The] challenge […] is to provide an account of the mechanisms that explain how our beliefs about these remote entities can so well reflect the facts about them. The idea is that *if it appears in principle impossible to explain this*, then that tends to undermine the belief in mathematical entities, *despite* whatever reason we might have for believing in them.[5]

What we are after, then, is some explanation of how it is that we—or at least mathematicians—are able to systematically form true mathematical beliefs. This issue arises for the mathematical Platonist because such a view rules out most straightforward mechanisms for explaining the correspondence between mathematical entities and mathematicians' beliefs about them.

While this challenge was first developed to target mathematical Platonism, there is nothing that makes it uniquely applicable to that domain. In fact, the

[3] Field (1989, 26). I gloss over certain caveats Field attaches to his formulation of the correlational claims.
[4] Field (1989, 25–26).
[5] Field (1989, 26), emphasis in the original.

requirement that the reliability of our beliefs must be explainable—that the correlation between our beliefs and the relevant set of facts requires an explanation—would seem to be quite general.[6] This type of requirement has later been dubbed the *reliability challenge*.[7]

This type of challenge should remind us of the motivation for Harman's explanatory constraint that was adopted by Joyce, discussed in §3.3.4. It should also remind us of Street's demand for an explanation of the overlap between the content of normative truths and that of our beliefs about them. This challenge can be captured by the principle UNEXPLAINED RELIABILITY, introduced in Chapter 4. Recall,

> UNEXPLAINED RELIABILITY
> If the reliability of S' belief that p cannot be explained, S' belief that p is undermined.

In the next subsection, I will discuss how to parse this rough formulation of the principle, including what reliability is, what it can be properly attributed to, as well was what it means to explain it. Before I do that, let us apply the challenge to the moral domain.

David Enoch has shown that the Benacerraf-Field challenge also arises for the normative domain, claiming that it is especially threatening to his own brand of normative non-naturalism.[8] Turning our attention to the moral domain specifically, the challenge is to explain the reliability of moral beliefs. Much like Field, Enoch takes this to require explaining the following correlation:

> [V]ery often, when we accept a [moral] judgment *j*, it is indeed true that *j*; and very often when we do not accept a [moral] judgment *j* (or at least when we reject it), it is indeed false that *j*.[9]

Before moving on to Enoch's suggested method for explaining this correlation—and thereby answering the reliability challenge—it is worth relating the reliability challenge to established ways of thinking about reliability and correlations. We will pursue that task in the next section.

[6] Cf. Enoch (2011, 176).

[7] As far as I can tell, the name originates with Schechter (2010), who argues that logical realism is subject to a reliability challenge of its own.

[8] Enoch (2011, 160).

[9] Enoch (2011, 159).

5.3 Reliability & Correlations

5.3.1 Explaining Reliability

To successfully answer the reliability challenge for the moral domain, one needs to explain the reliability of moral beliefs. What that task entails depends on what is meant by 'reliability.' This is not merely a semantic quibble, as 'reliability' is used with quite diverging senses in the philosophical literature.[10] First of all, reliability is attributed to different kinds of things in different contexts. In epistemology, reliability (or the lack thereof) is usually understood as an attribute of belief-forming methods, as in epistemological reliabilism.[11] In the literature on debunking arguments, however, reliability is often attributed instead to sets of beliefs.[12]

Second, whether attributed to sets of beliefs or belief-forming methods, reliability can have different modal scopes. Some take reliability to concern the truth or falsity only of *actual* beliefs. On such a view, reliability can be taken to consist in a particular proportion of actual beliefs (produced by the belief-forming method/in the set) being true.[13] Alternatively, reliability might be taken to signify that the actual beliefs (produced by the method/in the set) are *indicative* of truth.[14] It might also be taken to consist in the actual beliefs (produced by the belief-forming method/in the set) satisfying certain counterfactual conditions, such as epistemic sensitivity or safety.[15] Or it might be taken to be one of many other options.[16] Depending on which notion of reliability one operates with, what it takes to 'explain the reliability' of moral beliefs can vary.

Fortunately, for our purposes, we do not need to fully settle the issue of how to understand reliability. We can therefore to a large extent leave it open

[10] See Lutz (2020) for a discussion of the significance of different uses of the notions of reliability in debunking contexts.

[11] Goldman (1979).

[12] Enoch (2011, 155), Field (1989, 25–30), and Schechter (2010) take the issue to be, respectively, with the reliability of the set of our actual moral, mathematical, and logical beliefs. Tersman (2017, 758) understands reliability to be a feature of individual beliefs.

[13] Street (2006); Enoch (2011, 155).

[14] Armstrong (1973); Tersman (2017, 758). Tersman even applies it to individual beliefs.

[15] E.g. Clarke-Doane (2016).

[16] Other ways of understanding reliability include proper functioning (Plantinga 1993; Bergmann 2006) and epistemically virtuous belief-formation (Goldman and Beddor 2021, sec. 4.1).

exactly how one should understand the notion. I will nonetheless assume two things. First, anyone who is the proper target of the reliability challenge must accept *some* correlational claim of the type Field and Enoch set out. To avoid quibbling over whether a correlational claim, such as Enoch's and Field's, is too strong, we can instead employ a minimally demanding version of it. Call this the MINIMAL CORRELATION.

> MINIMAL CORRELATION
> At least some of the moral judgments I accept are true; and at least some of the moral judgments I reject are false.

Any non-skeptical realist or anti-realist *must* accept the MINIMAL CORRELATION.[17] If one denies the MINIMAL CORRELATION, any challenge springing from the need to explain the reliability of one's beliefs would be moot. Insofar as any proper target of the reliability challenge must accept the MINIMAL CORRELATION, and very likely an even stronger correlational claim, the reliability challenge requires that the correlation be explained.

One might take issue with the use of the word 'correlation' here. Consider an edge case where an agent claims to only accept *one* moral judgment that is true, and to reject only *one* judgment that is false. In that case, it would perhaps be improper to talk of a correlation between one's beliefs and the facts, strictly speaking. Even then, however, one would need to explain *how* one's belief *connects* or *relates* to the truth.

In such edge cases, therefore, we will be seeking an explanation of the *connection,* whether it is a correlation or not, between the content of one's beliefs and the facts. Such edge cases would serve my purposes equally well. However, most of us, and especially non-naturalists, will usually want to claim that we have more than one true positive and negative moral belief. Note also that the MINIMAL CORRELATION allows for there being a great many moral truths one does not know. It also allows other people, and even oneself, to have a great many false moral beliefs.[18]

[17] Cf. Field (1989, 26). For discussion of factors that can modify the strength of the correlational claim, see Enoch (2011, 166–67).

[18] Lutz (2020, 295) argues against the claim that some proportion of true beliefs is required for reliability. He claims that it is possible, at least in principle, to know that p, where p is about some domain D, despite having an overwhelming preponderance of false D-beliefs. This is not a problem for the MINIMAL CORRELATION, as it requires only a minimal number of true beliefs. Note also that in the cases considered by Lutz, the agent in question would *themselves* nonetheless assent to the MINIMAL CORRELATION and therefore be required to explain it.

By denying the MINIMAL CORRELATION, one would hold that there is *no* correspondence between one's own moral beliefs and the moral facts. In that case, one lacks any reason to believe that one's moral beliefs are reliable. This should be accepted on any plausible notion of 'reliability'. As discussed in §1.3, many, if not most, epistemic internalists, and many externalists would agree that holding one's beliefs to *in no way* correspond to the truth, serves to undermine those beliefs, for instance by being a defeater for them.

The reliability challenge, as formulated above, and the principle UNEXPLAINED RELIABILITY, raises further questions as well. Why is it that a failure to explain reliability has the power to undermine justification? There are many suggested explanations for this, but they tend to hold that the unexplained reliability of a set of beliefs (or of a belief-forming method) directly or indirectly constitutes a defeater for the beliefs in question.[19]

How to formulate the mechanism of undermining will depend on one's conception of justification. Consider a non-naturalist who accepts the MINIMAL CORRELATION, but who holds that there is no explanation of the correlation—no explanation of the systematic correspondence between her moral beliefs and the moral facts. Many hold that such *acknowledged* unexplained reliability is a sufficient condition for undermining the affected beliefs.[20]

It is worth noting that when it comes to *unreliability*, rather than unexplained reliability, internalists and externalists often allow that even *misleading* evidence of unreliability can undermine a belief. Goldman himself proposes a "no defeaters" clause for his reliabilism that allows for this.[21] One could argue analogously that even if a non-naturalist were to justifiably, but misleadingly, hold that there is no explanation of the MINIMAL CORRELATION, this could still be sufficient to undermine their moral beliefs.

One might think that on the above manner of spelling out reliability challenge, it has restricted scope. It might seem to require that for someone's moral beliefs to be undermined through the reliability challenge, they must *acknowledge* that there is no explanation of the MINIMAL CORRELATION. While this will hold true for many internalist frameworks of how defeasible justification is undermined, it is not necessarily the case.

One could hold that it is *the fact* that the reliability is unexplained that in turn explains why the beliefs are undermined. This could be accepted by externalists who do not accept the types of mental state defeaters discussed

[19] Enoch (2011, 161); Lutz (2020, 293–94). Alternatively, one could hold the weaker view that unexplained reliability is strong evidence against a view which needs to accept it (Baras 2017a).
[20] Enoch (2011, 160–61); Korman and Locke (2020).
[21] Goldman (1979).

above. In such cases, it would in general suffice that there is, in fact, no explanation of that correlation.

Another question is what models or mechanisms *are* available to the non-naturalist for explaining the MINIMAL CORRELATION. Consider, for instance, a moral naturalist who thinks that moral facts are reducible (without reminder) to, identical with, wholly constituted by, or fully grounded in natural facts. Such a moral naturalist need not think that moral facts are causally inert, and could therefore, at least in principle, provide some story about how our moral beliefs are causally explained by the relevant moral facts.[22]

A moral constructivist, on the other hand, denies that moral facts are stance-independent and could therefore, at least in principle, produce a story about how the moral truths are constitutively dependent on, and therefore constitutively explained by, our beliefs.[23]

The views that have been commonly thought to incur the steepest challenge from the reliability challenge are, as already mentioned, mathematical Platonism and moral (and normative) non-naturalism, as well as theistic Platonism, modal realism, and logical realism.[24] Common to all these views is a commitment to a domain consisting of facts and/or entities that are stance-independent, causally inert, and irreducible.[25] In virtue of these features, such views cannot rely on either of the above strategies for explaining their respective versions of the MINIMAL CORRELATION, such as causal mechanisms like perception or non-causal constitutive explanation.

The inability to appeal to causal or constitutive explanations is not in and of itself objectionable.[26] However, being blocked from appealing to such mechanisms could make it the case that there is no alternative explanation available. The lack of a causal or constitutive explanation therefore does not on its own constitute an objectionable feature, but it does make it incumbent on a defender of the relevant type of view to formulate what mechanism *could* explain the MINIMAL CORRELATION.

[22] For some such views, see Sturgeon (1985; 2006), Railton (1986), Boyd (1988), Brink (1989), Jackson (1998), and Finlay (2014). There might be other challenges remaining for such naturalist views, of course.

[23] Street (2008a) presents such a constructivist view.

[24] For discussion of the latter three in a debunking context, see respectively, Baras (2017b), Wang (2021), and Schechter (2010; 2013a; 2018b).

[25] These views differ in whether they introduce *entities* that are abstract and causally inefficacious, as mathematical Platonism does, but which moral realism plausibly does not.

[26] Cf. §2.2. Admittedly, Benacerraf's (1973) original formulation of the challenge did take the lack of a causal connection to be the crux of the issue, since he relied on the then-popular causal theory of knowledge. This is not the case in later refinements of the reliability challenge.

The non-naturalist could attempt to appeal to rational intuition in order to explain how we come to have reliably true moral beliefs. However, as Enoch argues, the mere appeal to such an ability does not help in this context.[27] This is because such an ability of rational intuition either operates through a causal mechanism, or it does not. If it does, it is incompatible with the non-naturalist's commitment to moral facts being non-causal. If not, it operates through a non-causal mechanism, in which case the question re-emerges of what this mechanism is and how it can explain the correlation.

Unless the non-naturalist puts forth a proposal for how the explanatory story is intended to go, appealing to such an ability is merely to label the explanation that is needed, but not provided. Having run out of ways to explain the correlation, the challenge goes, the moral non-naturalist would have to accept that moral beliefs are undermined by UNEXPLAINED RELIABILITY.

Of course, one way to explain the MINIMAL CORRELATION is to claim that it is the result of a lucky coincidence. There is something deeply unsatisfactory with this type of reply, however. An explanation of why it is unsatisfactory could be that coincidences can themselves undermine justification; it is widely agreed among epistemologists that coincidences can block true beliefs from being justified or from constituting knowledge.[28]

Take a simple example. Imagine that I formed beliefs about whether or not it will rain for each day of next month by guessing. Two weeks into next month, I have gotten everything right. In this situation, the explanation of the reliability of my weather beliefs will have to appeal to a large, patterned coincidence. This fact, in and of itself, would seem sufficient to undermine my beliefs about the weather outcomes for the remaining days.

Exactly how to capture the types of coincidences that are capable of undermining beliefs is a contentious issue that we will return to in the next chapter. For the purposes of this chapter, I will assume that attempting to explain the MINIMAL CORRELATION by claiming that the overlap of the content of one's beliefs with the content of the moral truths does not get a non-naturalist off the epistemological hook. In this chapter, I will therefore assume that something like NO COINCIDENCE, set out in Chapter 4, is true. Recall,

> NO COINCIDENCE
> For any theory, T, which concerns a domain, D, if T entails that the reliability of D-beliefs is merely coincidental, that strongly counts against T.

[27] Enoch (2011, 162); cf. Field (1989, 28).
[28] Pritchard (2007).

In the next chapter, we will look at proposals that attempt to explain what it is about coincidences that make them epistemically problematic, as well as at attempts to reject NO COINCIDENCE.

Let us take stock so far. The reliability challenge is answered by providing an explanation of the reliability of moral beliefs. One necessary component of explaining this reliability is to explain the MINIMAL CORRELATION. Some debunkers have thought that this requirement leaves only one possibility for the non-naturalist—to appeal to a coincidence.[29] If some principle like NO COINCIDENCE is true, this leaves the non-naturalist without a reply to the reliability challenge.

In response to this predicament, there has been a concentrated effort to come up with an explanation of the correlation that does not appeal to an epistemically problematic coincidence. Before moving on to consider one such attempted explanation, it will be instructive to take a closer look at how correlations can be explained.

5.3.2 Explaining Correlations

Correlations are systematic associations between phenomena. Smoking and lung cancer are correlated; an increase in smoking frequency is associated with an increase in the likelihood of contracting lung cancer. As we now know, this correlation is explained by the fact that smoking *causes* lung cancer. This illustrates the first method for explaining a correlation between two variables—establishing that there is a *direct* relationship between them. In this case, furthermore, the direct relationship is a causal one.

As every introductory statistics textbook emphasizes, two variables being correlated does not imply that there is any causal relationship between them. The variables could stand in a direct, non-causal relationship. For instance, they could do so by co-occurring, as in the case of being human and having a human parent. Or it could be that one variable is constitutively related to the other, as in the correlation between soccer balls passing the goal line during a match and a goal being scored.

Two correlated variables need not have any direct relationship at all. This is because there could be some additional variable—a so-called third factor—that *indirectly* explains the correlation between the two initial variables. The

[29] Bedke (2009; 2014) sharpens Street’s argument and develops the idea that the coincidence a non-naturalists would need to embrace is a defeater for the relevant beliefs.

third factor could do so by itself being causally connected to the two initially correlated variables or by explaining them in some non-causal manner. I will give an example of both types of third-factor explanation.

As a case where the third factor is causally responsible for both the initially correlated variables, consider two variables in a population of school children—height and vocabulary.[30] These variables are correlated; greater height is associated with a larger vocabulary. In this case, it is implausible to hold that there is some direct connection, whether causal or non-causal, between height and vocabulary size. An alternative explanation of the correlation is that what explains both these variables is a third factor, such as physio-cognitive development. Such development is capable of causally explaining both why a school child's height and vocabulary tend to increase over time. This shows how there can be an indirect causal explanation of the two initially correlated variables, by way of a third factor, without the initially correlated variables themselves being directly related.

Consider now a different interpretation of the same case, which illustrates how third-factor explanations can be non-causal, but explanatory nonetheless. An explanation of the correlation between height and vocabulary size among school children can appeal to a different third factor, such as grade level. A school child's grade level does not causally influence their height nor their vocabulary size but is itself correlated with both. As school children advance in grade levels, they tend to experience an increase in vocabulary size and also tend to see an increase in height.

Since this is not a causal explanation, but merely an appeal to another pair of correlations—that between grade level and height and between grade level and vocabulary size—one could wonder if any explanatory progress has really been made by appealing to them. To see that it has, one can control for the third factor's explanatory contribution by holding grade level fixed (i.e., only consider students in the same grade level) and checking whether there is nonetheless a systematic association between height and vocabulary. If there is still such an association, the fixed factor did not fully explain the correlation. However, if the association disappears, the fixed factor is likely to have provided some explanatory power.[31]

In the above example, holding grade level fixed would presumably eliminate, or at least greatly reduce, any correlation between height and vocabulary

[30] I borrow this example from Warner (2013).

[31] This is a highly simplified way of describing the process of screening off factors. See e.g. Warner (2013, chap. 10) details.

size.[32] In this way, grade level serves to explain the correlation between the two initial variables in an indirect, non-causal manner; one is usually taller in higher grade levels, and one's vocabulary tends to increase as one's grade level increases. The original correlation is now explained in terms of two further correlations—that between the third factor and each of the initial variables separately. Appealing to a third factor, whether causally or non-causally related to the original variables, constitutes a second method in which a correlation can be explained.

There is also a third way of explaining correlations, which covers what is sometimes called *accidental correlations*.[33] These are correlations where there is no direct relationship between the correlated variables, and neither is there any third factor that can provide an indirect explanation of the correlation. The systematic association between the two phenomena is therefore merely brute or accidental.

One such purported correlation is that between British bread prices and Venetian sea levels.[34] This correlation can only be explained, to the extent that it can be explained at all, by explaining the occurrence of each variable separately. Unlike the previous forms of explanation considered, explanations appealing to an accidental correlation are therefore not *unified*. That is, there is no explanatory relation that holds between the associated variables, either directly or indirectly. In the phrase of one statistics encyclopedia, such systematic associations are 'cosmic coincidences.'[35]

These three forms of explanation—direct, indirect, and accidental—are the ones commonly accepted in fields traditionally employing correlational studies and data, such as social science.[36] In addition to these, more rarified ways

[32] A reason why this might not fully eliminate the correlation is that children who are born early in the year, and are therefore older than their classmates, are likely to both have a larger vocabulary and be taller.

[33] They are also sometimes called *nonsense correlations*, although this term is sometimes used to include the type of correlations described in fn. 36 below, which results from methodological issues.

[34] Sober (2001).

[35] Haig (2007, 938). As it happens, this term has also been used in the debunking literature in similar contexts (Bedke 2009, 197).

[36] Haig (2007), who provides a similar but slightly different taxonomy. He adds a fourth category, which are not actual correlations but rather "artifacts of method and arise from factors such as sample selection bias; use of an inappropriate correlation coefficient; large sample size; or errors of sampling, measurement, and computation" (2007, 938). These are therefore not really correlations after all, and so they can more properly said to be explained *away*.

of explaining correlations have been developed in philosophical and mathematical contexts. For instance, in response to the Benacerraf-Field challenge, some have suggested adopting so-called *full-blooded Platonism*.

This is the view that some domains are very richly populated, ontologically speaking, such that *any* theory of that domain that satisfies certain criteria (e.g., consistency) is true.[37] This is because, according to this view, some domains are *plenitudinous*. This view has been applied to the mathematical domain, where to say that the mathematical domain is plenitudinous, is, roughly, to say that

> all consistent purely mathematical theories truly describe some collection of abstract mathematical objects. Thus, to acquire knowledge of mathematical objects, all we need to do is acquire knowledge that some purely mathematical theory is consistent.[38]

The explanation of how we come to have reliable mathematical beliefs, on this view, is that our mathematical theory (if consistent) *could not have failed to be true*. It is important to note that our belief in a consistent mathematical theory does not *determine* or *generate* mathematical reality. Instead, those beliefs merely determine *which part* of the plenitudinous mathematical reality is picked out by our beliefs. If an analog of this view were available to the moral non-naturalist, they would have a further model for explaining the MINIMAL CORRELATION.

If a non-naturalist is to explain the MINIMAL CORRELATION, she will have to opt for one of the four types of explanations just outlined—direct, indirect, accidental, or full-blooded Platonism. Since the moral non-naturalist denies that moral facts and properties have causal powers, they cannot hold that the moral facts explain our beliefs by being directly causally related to them. Since they hold that moral facts are stance-independent, they cannot hold that our beliefs constitute or otherwise directly explain the moral facts. This therefore rules out the first type of explanation—a direct relationship in either direction. This leaves three options, either appealing to an indirect explanation in terms of a third factor, taking the correlation to be accidental, or accepting a plenitudinous moral reality. I will quickly consider these in reverse order.

Imagine first that a non-naturalist postulates a plenitudinous moral reality, where any moral theory that satisfies certain constraints, such as consistency, is true. This brings with it certain puzzling consequences. Most importantly,

[37] Balaguer (1998).

[38] Balaguer (1998, 48).

as Justin Clarke-Doane has argued, it leads to the loss of a uniquely right answer to many important moral and practical issues.[39]

Assume for instance that both some version of deontology and consequentialism are consistent and that this is all that is needed for picking out some genuine part of a plenitudinous moral reality. Since both consequentialism and deontology make pronouncements about what one ought to do, the following scenario is likely to arise. For some action φ, it will be the case that you ought to φ, when evaluated relative to a consequentialist framework. When evaluated according to a deontological framework, however, it is not the case that you ought to φ. Since there is, presumably, no metaphysically privileged part of the plenitudinous moral reality, moral theories cannot provide any guidance to the practical question of what to do, without evaluating it relative to a particular moral framework. That is, moral theory can then provide no unique answer to the deliberative question agents often ask themselves: “Taking every moral consideration into account, how do I act?”

This is similar to how a mathematical question, such as whether Euclid’s parallel postulate is true *simpliciter* arguably has no answer. There are distinct, consistent, mathematical theories—e.g., Euclidian and non-Euclidian geometries—which give conflicting answers. While it is not an unacceptable consequence for a theory of mathematics to claim that there are no metaphysically privileged answers to certain mathematical questions, many would take the inability to provide a uniquely true answer to moral and practical issues to be a *reductio* of a realist metaethical view.[40] This is especially true for the non-naturalist. An explanation of the MINIMAL CORRELATION modeled on full-blooded Platonism is therefore not a good fit for the non-naturalist.

Opting instead for an accidental explanation would be to appeal to a brute coincidence. Since we are for now assuming NO COINCIDENCE to be true, that is not a viable option. It might therefore seem prudent to explore the possibility of an indirect, third-factor explanation. This is just what many non-naturalists have done.

[39] Clarke-Doane (2020, 164).

[40] Admittedly, some have argued that similar issues haunt other metaethical theories beyond the moral version of full-blooded Platonism, non-naturalism included (Eklund 2017; 2020). However, for other views, this consequence is at the very least a bug and not a feature.

5.4 Third-Factor Explanations of Moral Beliefs

Let us now consider the attempt to answer the reliability challenge by providing a third-factor explanation of the reliability of moral beliefs. Call this—employing a third-factor explanation to answer the reliability challenge—*the third-factor strategy*.[41]

Having set out his version of the reliability challenge, Enoch suggests the following course of action for non-naturalists.

> The thing to look for is a third-factor explanation. For it is possible that the explanation of a correlation between the two factors A and B is in terms of a third factor, C, that is (roughly speaking) responsible both for A-facts and for B-facts.[42]

The third-factor strategy seeks to combine two elements that, in Street's Darwinian Dilemma, supposedly lead to skepticism: subscribing to stance-independent moral facts while simultaneously holding there to be no direct causal or constitutive explanatory connection between the moral facts and our moral beliefs. The third-factor strategy circumvents Street's dilemma by letting a third factor play a dual explanatory role, helping to explain both moral beliefs and the moral truths, without positing any direct explanatory relation between the two.

If we apply the models of third-factor explanations set out in the previous section, there are two ways in which a non-naturalist could proceed. First, they could provide an account where the third factor is merely correlated with each of the two initial variables. Call these *indirect third-factor explanations*. Second, they could opt for the type of third-factor explanation where the third factor itself stands in a direct explanatory relationship with each of the associated variables—moral beliefs and moral facts. Call these *direct third-factor explanations*.

Let us now consider some examples of third-factor explanations from the literature and see how they relate to this distinction. In an overview of work on evolutionary debunking arguments and third-factor explanations, Erik Wielenberg sets out the issue as follows.

[41] For some non-naturalist defenses of third-factor strategies, see Enoch (2010; 2011, chap. 7) and Wielenberg (2014, chap. 4). Naturalists have also developed third-factor explanations, or something very much like it, of the reliability of moral beliefs (Brosnan 2011).

[42] Enoch (2011, 168).

> One popular sort of response to [evolutionary debunking arguments] is to propose some factor (a so-called 'third factor') that is correlated with both our moral beliefs and the moral truths and so plausibly explains how our moral judgements could be correlated with the moral truths even if the moral truths do not even partially explain our moral beliefs.[43]

Wielenberg here chooses to focus on the third factor being *correlated* with, rather than directly explanatorily connected to, the moral truths and moral beliefs. Wielenberg then goes on to introduce one such purported third-factor explanation.

> For example, David Enoch proposes that survival or reproductive success is good. If that's so, then since evolutionary forces have pushed us towards moral judgements that cohere with the goodness of reproductive success, it turns out that evolutionary forces have generally pushed us in the direction of the moral truth.[44]

Wielenberg is not explicit about what the third factor actually is, but seems to be suggesting that it is "that survival or reproductive success is good." This is not surprising, as Enoch himself often frames his account this way.

> The connection between evolutionary forces and value—the fact that survival is good—is what explains the correlation between the response-independent normative truths and our selected-for normative beliefs.[45]

Similarly, many other purported third-factor accounts are framed as if the third factor is one or more substantial normative or moral facts.[46] Let us therefore consider the suggestion that it is a singular normative fact—that survival is good—that constitutes the third factor, and that it is correlated, rather than directly explanatorily connected to, both the moral facts and the moral beliefs.

What does it mean to claim that the singular normative fact that survival is good is *correlated* with our moral beliefs? Or that this singular normative fact is *correlated* with the moral facts? It is not altogether easy to see how to spell out such claims, especially not in any way that would help the non-naturalist. In fact, I can see no plausible way of making sense of the claim that this sin-

[43] Wielenberg (2016, 505).
[44] Wielenberg (2016, 505–6).
[45] Enoch (2011, 168).
[46] For instance, Skarsaune (2011) is often claimed to defend a third-factor explanation, where the third factor is purportedly that "Pleasure is usually good and pain is usually bad" (2011, 232). Skarsaune does not himself use this label, and as we will see, this is for a good reason.

gular normative fact can function as a third factor in the sense of being correlated with both moral beliefs and the moral facts. As I will discuss below, there are, however, other ways to make sense of such claims.

This rejection of singular normative facts as third factors might seem premature and uncharitable. To address this suspicion, I believe the best course of action is to provide an interpretation of claims similar to those of Enoch and Wielenberg which vindicates them, but which removes their status as a third-factor explanation. I will argue that this makes better sense of the relevant claims.

Consider again the large amount of literature suggesting a conception of the third factors occurring in third-factor explanations like the following:

> It is important that in [Enoch's and Skarsaune's] arguments, the third factors (Survival is good, Pleasure is good) […] are themselves members of the set of alleged mind-independent normative truths that the realist is attempting to defend in light of Street's conclusion that our moral judgments are unlikely to be largely accurate reports of mind-independent moral truths.[47]

> The third factor in Enoch's third-factor explanation is the truth of a normative claim.[48]

On such a construal, the singular normative fact that survival is good is intended to explain the relevant correlation, but how? Here is a sensible proposal, which is sometimes mirrored in Enoch and the surrounding literature. First, there is the singular normative fact that survival is good. This fact stands in various relations to other normative facts. For instance, if survival is *pro tanto* good, then my survival is *pro tanto* good. On the other hand, because of the purported evolutionary advantages of survival, evolutionary pressures will lead us to believe that survival is good. That is, we are disposed to form beliefs with that content. This means that evolutionary processes, because of purely biological mechanisms of promoting fitness, were likely to lead us to have the belief that survival is good.

The normative facts being what they are—e.g., survival being good—makes sure that there is an overlap between the content of the normative facts and the content of the beliefs evolutionary pressures dispose us towards. In this way, the fact that survival is good at least partly helps explain the overlap of the content of moral beliefs and the moral facts.

[47] Dyke (2020, 2119). Dyke also discusses a third-factor explanation due to Wielenberg (2010), but I set that aside here.

[48] Behrends (2013, 492).

I believe the above is a charitable interpretation of what many intend to defend as being a third-factor explanation. I now want to suggest that *this* type of explanation involves no third factor and is therefore not a third-factor explanation at all. Recall that the defining feature of a third factor is that it plays a *dual* explanatory role; it is either directly or indirectly explanatorily connected to *both* the variables whose association it is intended to explain.

Consider again the fact that survival is good. This fact could conceivably be explanatorily related to other moral facts, for instance by standing in various relations to them, such as coherence or constitution relations. However, this normative fact in no way—either directly or indirectly—explain our moral beliefs. *Ex hypothesi,* it is the causal evolutionary influences on our beliefs (which inclines us towards having the belief that promoting survival is good) that explain them.

That this is so can be seen by doing a simple screening test. Bracket the first *relatum* (the normative fact that survival is good), and the causal evolutionary explanation of our moral beliefs is no worse off for it. We would, *ex hypothesi*, still have the same moral beliefs that we in fact do. Bracket the second *relatum* (the evolutionary influence on our beliefs), and the explanation of the moral facts (by constitution or coherence) loses no power. It is still the case that, say, if murder is always wrong, then murdering your neighbor is wrong.

If we had made explanatory progress in explaining the correlation by proposing that survival is good is a third factor, this would not have been the case. In light of this, there is simply no way of making sense of the claim that a singular normative fact—such as survival being good—could directly or indirectly explain, be responsible for or otherwise stand in the required relation to the fact that we have the moral beliefs that we do. This, in turn, is because there is no direct or indirect explanatory connection between the singular normative fact and our moral beliefs. I therefore submit that the above type of explanation, which employs a singular normative fact, is not a third-factor explanation at all.

To see this more clearly, consider the structure of the explanation provided by the claim that "survival is good pre-establishes the harmony between the normative truths and our normative beliefs." This explanation should not be understood as offering *one* factor that explains the association of two correlated variables, A and B. Instead, it should be understood as offering a conjunction of the separate and independent explanations of A and B. The singular normative fact helps explain other normative facts, while the evolutionary explanations of belief, explain moral beliefs. Luckily, the content of the facts and of the beliefs overlap because of the general story presented above. There is, in other words, no single factor here that plays a dual explanatory role with respect to *both* A and B.

As we saw in the previous subsection, when a correlation between two variables is explained by an appeal to two separate, non-unified, independent explanations of the associated variables, what you get is an *accidental correlation*. As we saw in §5.3.2, an accidental correlation is one where there is no direct relationship between the associated variables, and neither is there some indirect relation—i.e., a third factor—connecting them. Instead, the association of the variables is something of a cosmic coincidence. Such an explanation of the MINIMAL CORRELATION—appealing to a cosmic coincidence between moral beliefs and moral facts—might conceivably deserve the moniker "pre-established harmony", but it is not an instance of a third-factor explanation.

For the reasons set out above, I think substantial parts of the literature on third-factor explanations have not, in fact, been concerned with such explanations at all. Instead, they have been concerned with discussing and defending what I will call the *accidental correlation strategy*.[49] We will consider this strategy in the next chapter, so I will set it aside for now. Despite third factors regularly being identified with normative facts in the literature, many formulations of third-factor accounts do not make this mistake.

Having now said something about what a third-factor explanation is *not*, and which form it is unlikely to take, it is time to give an example of what a third-factor explanation could look like. Below, I will set out a modified version of Enoch's proposal which adheres to the structure of direct third-factor explanations set out in §5.3.2. Later in this chapter, I will present two objections to the third-factor strategy that apply to all such third-factor explanations. Before that, it will be helpful to first illustrate the strategy with a particular third-factor proposal.

5.4.1 An Alternative Version of Enoch's Third-Factor Proposal

I will now present an interpretation of Enoch's general proposal which does in fact qualify as a third-factor explanation. This interpretation, based on work

[49] I believe this confusion largely stems from Enoch's (2011) formulations, which can more naturally be read along the lines of the accidental correlation strategy, and only with some charitable interpretation as a third-factor explanation. I will provide the latter reading below.

by Folke Tersman, diverges from Enoch's own formulations.[50] For instance, while Enoch focuses on evaluative properties such as *goodness*, I will extend the consideration to deontic properties such as *rightness*.

The interpretation third-factor explanations such as Enoch's that I am concerned with consists of two main elements. First, there is a normative assumption—a bridge principle—which for Enoch is that "survival is good."[51] The second element of Enoch's proposal on this interpretation is the empirical claims, which are meant to show that our moral beliefs have been shaped by evolutionary forces in a way that aligns them with the normative truths laid out by the bridge principle.

Enoch does not provide much in the way of specifics on this latter point but keeps his talk abstract and metaphorical: "Selective forces have shaped our normative judgments and beliefs, with the 'aim' of survival or reproductive success in mind (so to speak)."[52] What is needed from the empirical claims is to make it plausible that our moral belief-forming processes are subject to significant evolutionary pressures along the lines familiar from debunking arguments such as those presented in Part I.

With these two elements in place, what is the third factor that is supposed to help explain, at least to some extent and in some respect, *both* the moral facts and the moral beliefs? As we have seen, this is often less than clear both in Enoch's writing and in much of the secondary literature. To see what the third factor could be, let us consider how moral beliefs and moral facts could be explained. Enoch does not flesh this out in any detail, but we can elaborate slightly. Start by noticing that, plausibly, selective pressures have influenced our moral beliefs in a way that inclines us towards having pro-attitudes towards various fitness-enhancing behaviors and states of affairs. This is because having such pro-attitudes tends to increase reproductive success.

Take the belief that it is morally right to take care of one's children. How could we explain the widespread distribution of this belief in the human population? By appealing to the fact that having this belief tends to help promote the survival of one's children, and is therefore fitness-enhancing. Similarly, other moral beliefs that the non-naturalist would wish to vindicate the reliability of would have to be explained by appealing to their role in promoting

[50] Tersman (2017).

[51] Enoch (2011, 168). In an even weaker formulation, he says that it can be understood as "survival (or whatever) is actually by and large better than the alternative" (2011, 168).

[52] Enoch (2011: 168).

survival, or by some other indirect method such as coherence relations with the aforementioned fitness-enhancing beliefs.[53]

On this fleshed-out version of Enoch's proposal, a factor that helps explain why we have particular moral beliefs that ascribe rightness to acts is *the facts about which actions promote survival*. These facts about survival promotion are part of the explanation of the causal origins of our moral beliefs.

Consider now how moral facts could be explained on this proposal. We are assuming that survival is (at least somewhat) good. Taking care of one's children promotes their survival, and hence it promotes the good. If we allow that an action promoting the good contributes to the rightness of that action, then taking care of one's children is to that extent right.[54] So, the rightness of an action is party explained by the fact that it promotes survival. This means that one factor that helps explain which actions are right is *the facts about which actions promote survival*.

As we can now see, one factor that helps explain *both* our beliefs about which actions are right and the facts about which actions are right is *the facts about which actions promote survival*. This set of facts, or something close to it, is a prime candidate for being the third factor in this modified version of Enoch's proposal. The bridge principle—that survival is good—is what helps bridge the rightness-facts (e.g., that it is right to take care of one's children) with the content of the moral beliefs that result from evolutionary influences (e.g., the belief that it is right to take care of one's children). The bridge principle therefore allows the natural facts about survival-promoting behavior to play the dual explanatory role needed of a third factor. Even so, the bridge principle itself is *not* the third factor, as it is often claimed to be.

It is worth noting that the third factor does not necessarily do *all* the explanatory work with respect to either the moral beliefs or the moral facts. But that is not a requirement for a third-factor explanation. What is required is that there is some factor that plays an appropriate explanatory role in the explanation of both the associated phenomena—in this case, the moral beliefs and the moral facts. At first glance, the above interpretation of Enoch's proposal might seem, at least in principle, to be able to satisfy that requirement.

Any specific proposal, such as the above, might turn out to be a bad one on account of its content, and might eventually be rejected on that basis. But before considering the plausibility of specific proposals, it is worth considering

[53] Enoch (2011, 166). Cf. FitzPatrick (2015, 893–94).

[54] Enoch (2011, 169) himself endorses something like this, though his discussion proceeds in terms of the goodness of acting in certain ways rather than the rightness of acts.

whether the structure of the third-factor strategy is such that it could, in principle, deliver an explanation of the MINIMAL CORRELATION.

Coming to think that a third-factor explanation could, at least in principle, succeed in explaining the MINIMAL CORRELATION would go some way towards staving off the threat from the reliability challenge. Conversely, if third-factor explanations of the reliability of moral beliefs seem even in principle unworkable, then there is not much need to consider the plausibility of particular proposals.

I am now going to consider two objections to the third-factor strategy that focuses on general, principled features of the strategy. First, I will discuss a very conspicuous feature of the strategy that many have taken issue with: the reliance on substantial normative or metaethical assumptions as bridge principles. I will ultimately set it aside, as I find that it results in a legitimate dialectical stalemate between a debunker and the non-naturalist. I then go on to provide a novel objection to third-factor explanations which shows that the third-factor strategy, when properly understood, commits its proponents to a picture of moral explanation that makes it impossible for non-naturalists to employ it.

5.5 Is the Third-Factor Strategy Question-Begging?

Many have felt that the third-factor strategy, in Russell's phrase, "has many advantages; they are the same as the advantages of theft over honest toil."[55] In particular, many have been frustrated by the fact that a strategy that intends to vindicate our beliefs about moral facts—in the face of skeptical arguments—relies on substantial normative assumptions concerning *those very same facts*. Street herself voiced an early version of this complaint:

> It is no answer to [the Darwinian Dilemma] simply to assume a large swath of substantive views on how we have reason to live […] and then note that these are the very views evolutionary forces pushed us toward. Such an account merely trivially reasserts the coincidence between the independent normative truth and what the evolutionary causes pushed us to think; it does nothing to explain that coincidence.[56]

[55] Russell (1919, 71). Russell was concerned, not with debunking arguments, but with methods in the philosophy of mathematics.

[56] Street (2008b, 214) in response to Copp (2008).

It is, at least at first blush, easy to have sympathy with such an objection. Making substantial moral assumptions about what is good or right in the course of arguing that our moral beliefs are reliable, has seemed plainly and objectionably question-begging to many.[57] I will consider two ways in which this objection can be made more concrete.[58] I will argue that while the objections might seem superficially appealing, there is no straightforward way of spelling them out such that they succeed in blocking the third-factor strategy.

5.5.1 Relying on Contested Evidence

The third-factor strategy relies on a normative assumption—the bridge principle. For a particular proposal to be plausible, the bridge principle it employs must itself be plausible. When attempting to show that the bridge principle is plausible, the third-factor strategist relies on our antecedent moral beliefs and intuitions to defend it.[59] Establishing the plausibility of the bridge principle therefore requires introducing as evidence *exactly* the evidence that is being contested by a debunker (i.e., the reliability of our moral beliefs).

In its most general from this objection relies on a principle that, in a given dialectical context, bans the introduction of any evidence that is contested in that context. Such a ban would require that any evidence introduced in a given dialectical context be neutral with respect to the issue at hand.[60]

Despite the intuitive force of such a requirement that evidence be neutral between disputing parties, it is not at all clear that it can ultimately be defended. If such a ban were in place, one would count as having evidence for one's view only as long as that evidence is not contested by one's opponent. In a dialectical setting where one's opponent contests much, or even all, of one's evidence—as a global skeptic would—one will be left without evidence for one's view.

When faced with a large enough skeptical challenge in a given dialectical context, one would therefore succumb to it.[61] Clearing the road for wide-ranging skepticism, even if localized to a particular context, is not an attractive

[57] Street (2008b, 214; 2011, 17–21); Shafer-Landau (2012, sec. 6); Fraser (2014, 471); Vavova (2014, 81–82); Crow (2016, 289–91); Joyce (2016a, 157–58); Lott (2018, sec. 2.2).
[58] See Korman and Locke (2020, 314–16) for another attempt at precisification of the charge of question-beggingness.
[59] As Enoch (2011, 170–71) himself does.
[60] For discussions of such a requirement of evidential neutrality, see Nozick (1981, 197–98), Pryor (2004, 368–70), Williamson (2007, 220–25), and Kelly (2008, 73–76).
[61] Williamson (2007, 238).

result for most debunkers.[62] They tend to seek targeted challenges for one, or at most a few select, domains, such as morality or normativity generally.[63]

If, in response, the debunker relaxes the requirement that evidence be neutral between the disputing parties, it is no longer necessarily illegitimate to rely on evidence that your opponent does not accept. It could then be legitimate to introduce evidence in defense of one's view, despite it not being of the kind that would persuade an interlocutor in a debunking context, or even be such that the interlocuter accepts that it *is* evidence.

These are contested and complicated issues. But without refining the objection in some further way, it does not seem to stick. If there is a problem with the third-factor strategy, it cannot merely be that it relies on evidence that is dialectically contested by a debunker.

It is worth nothing that this way of formulating the worry about the strategy being question-begging can help explain the impression of a dialectical stalemate that seems to surround third-factor strategies in debunking contexts. Proponents put them forth and propose bridge principles they find sensible. Debunkers, on the other hand, find such bridge principles question-begging and therefore reject the proposals along with the strategy. Without a ban on contested evidence, this stalemate could be entirely legitimate. It would therefore behoove debunkers to find another way of framing the objection if they wish to secure dialectical traction with the non-naturalist.

5.5.2 Circular Establishment of Process Reliability

A second way to spell out the objection that third-factor explanations are question-begging, is to claim that they allow for a circular vindication of the reliability of the belief-forming processes generating moral beliefs. Consider a version of this objection coming from Russ Shafer-Landau.

> [W]e are worried about the reliability of our moral faculties. For all we know, they might be drastically unreliable. It would be illicit at that point to introduce moral beliefs to bolster our confidence in our moral faculties, as any doubts about the faculties themselves ought to be transmitted to the beliefs they generate.[64]

[62] Vavova (2014).
[63] E.g. Street (2016, 319).
[64] Shafer-Landau (2012, 33).

The reliability challenge requires that the non-naturalist explain the reliability of moral beliefs. For this purpose, the non-naturalist claims that it is acceptable to enroll the outputs of the processes that underlie the formation of moral beliefs in order to establish their reliability. One might feel, with Shafer-Landau, that this is plainly circular and question-begging. This is not least because it would allow us to establish the reliability of a belief-forming process by *assuming that the process is reliable*. Richard Fumerton expresses the frustration a skeptic might feel in the face of this strategy.

> You cannot *use* perception to justify the reliability of perception! You cannot *use* memory to justify the reliability of memory! You cannot *use* induction to justify the reliability of induction! Such attempts to respond to the skeptic's concerns involve blatant, indeed pathetic, circularity.[65]

While appealing to the output of a belief-forming process in order to establish the reliability of that same process might indeed seem ‘pathetically circular’, denying the legitimacy of such a move comes at a cost. Take any source of beliefs that we would want to establish the reliability of, such as perception. Consider now some skeptical hypothesis that challenges the reliability of perception. When tasked with providing evidence that perception is reliable *without ever appealing to the outputs of that process*, one would be at a loss. When widening the skeptical charge to *all* belief-forming processes, there would be no possibility of vindicating the reliability of *any* beliefs. Adopting a fully general prohibition on the circular establishment of process reliability therefore leads to global skepticism.[66]

If a debunker opts for defending a ban on this way of establishing reliability, they are again led to push a much wider skeptical challenge than they usually intend. This will not be attractive to most debunkers who seek to challenge certain select domains of belief and not our general reliability.

A possible way out of this predicament is suggested by Korman and Locke.[67] They propose that a debunker could grant, at least with respect to certain sources of belief, that we are entitled to trust their outputs *by default*. The question then becomes exactly *which* sources of belief have this feature of granting default entitlement. This is up for debate, but natural candidates include perception, introspection, memory, and testimony. For this response from the debunker to succeed, it would be necessary to show that such default entitlement *does not* extend to moral beliefs.

[65] Fumerton (1995, 177), emphasis in the original.

[66] Alexander (2011).

[67] Korman and Locke (2020).

While it is certainly open to debunkers to argue that morality should be excluded from any list of potential sources of default entitlement, doing so would require a separate argument. A worry about this move is that any such separate argument that is sufficient for denying default entitlement to moral beliefs would likely involve showing that moral beliefs are epistemically problematic, something which risks making the reliability challenge redundant.

There is also a worry about what such a move would mean for epistemic beliefs. The debunker employs epistemic premises in their debunking argument, and unless the reliability of the belief-forming processes generating epistemic beliefs are either granted default entitlement, or can be vindicated by processes that are, the argument would face the threat of self-defeat.

However this may be, there would seem to be no straightforward route to introducing a general ban on circular methods of establishing the reliability of sources of belief, if one does not want to embrace wholesale skepticism.

The primary problem with holding the third-factor strategy to be question-begging, as we have now seen, is that any epistemological principle sufficient to underwrite the objection leads to a much broader form of skepticism than most debunkers intend to argue for. The debunker therefore faces a choice, either to plump for reframing moral debunking arguments as an instance of a familiar type of general skeptical challenge or withdraw the objection that the third-factor strategy is problematically question-begging. The former option should seem unattractive to the debunker, as it will make debunking arguments of the type considered in this thesis (specifically against morality or normativity) uninteresting, since they will simply be, as Street agrees, a "routine, general skeptical worry deployed selectively."[68] Few debunkers seek to push such a broad debunking challenge.[69]

For these reasons, I am not convinced that worries about question-beggingness will, in the end, pose a significant challenge for the third-factor strategy. In any case, I will not push them further here. Instead, I will develop an objection against the third-factor strategy that will have force even if worries about question-beggingness can be assuaged.

[68] Street (2016, 319).

[69] A debunker who at least might appear, in some moods, to opt for such a broader debunking argument is Ruse (1986).

5.6 The Structure of Moral Explanations

If the third-factor strategy is not objectionably question-begging, then that no longer constitutes grounds for a principled objection against it. In this section, I will develop a different principled objection against the strategy. This objection arises from the metaphysical commitments of non-naturalism, which I will claim are incompatible with the requirements of a third-factor explanation of moral beliefs. My central claim will be that, given an attractive, natural, and perhaps necessary picture of the structure of moral explanation, it becomes impossible for non-naturalists to find a candidate third factor that could, even in principle, play an explanatory role with respect to both fundamental moral facts and our beliefs about them. Consequently, non-naturalists cannot successfully employ the third-factor strategy to explain the reliability of moral beliefs.

Recall what it would take to provide a successful third-factor explanation of the reliability of moral beliefs. Any third factor to earn the name would have to be the kind of thing that could explain, whether causally or otherwise, and at least in part, both the moral facts and our beliefs about them. To achieve this, the third factor must be responsive to facts about things like survival or evolutionary fitness since, *ex hypothesi*, those are the facts that explain our moral beliefs. As we saw with the revised version of Enoch's proposal, they could be natural facts concerning the effects of certain action types.

In addition to explaining our beliefs, a third factor must be capable of playing an explanatory role with respect to moral facts, including the fundamental ones. A factor that could play such a dual explanatory role would be a suitable fit for a third-factor explanation of the reliability of moral beliefs.

As we have seen, Enoch and many others are optimistic about the project of locating such a third factor. I will now argue that this optimism is misplaced, at least for a non-naturalist.

Consider UTILITARIANISM.

> UTILITARIANISM
> An action is right if and only if, and because, it maximizes utility.

According to a utilitarian, UTILITARIANISM is not only a moral fact, but a fundamental moral principle. Any candidate third factor should help explain it. There are at least three reasons why satisfying this expectation will not be possible. The first reason is that non-naturalists, with few exceptions, hold that fundamental moral principles are metaphysically necessary.[70] Insofar as natural facts are unlikely candidates for explaining (or otherwise being responsible for) metaphysically necessary facts, the type of third factors we have considered above—natural facts about action types—seem ill-suited for the role. Let me illustrate.

If UTILITARIANISM is true, then it seems that every non-fundamental moral fact will, in part, be explained by the natural fact that a given action maximizes (or fails to maximize) utility.[71] If it's morally right to give 10% of one's income to charity, then that fact is partly explained by the natural fact that performing the act would maximize utility.

But what explains the fundamental moral facts—i.e., UTILITARIANISM? This depends on one's view of the structure of moral explanation. On one such view, a principle like UTILITARIANISM is a universal generalization that is fully explained by its instances.[72] The principle itself would then not play a role in explaining any non-fundamental moral facts, but is explanatory merely in virtue of describing or capturing a kind of regularity (i.e., the regularity that utility-maximizing actions are right). Whatever the plausibility of such a view, it is not a good fit for a non-naturalist, for reasons to be discussed below.[73]

On an alternative view, more friendly to the non-naturalist, moral principles are *not* fully explained by their instances and *do* play an explanatory role with respect to non-fundamental moral facts. On such a view, the fact that it is morally right to give 10% of one's income to charity is partly explained by the natural fact that doing so would maximize utility *and* partly by UTILITARIANISM.[74]

On this view of moral explanation, the fact that any particular action, or set of actions, maximize utility, plays no part in explaining the truth of UTILITARIANISM. The type of proposed third factors discussed above—involving natural facts—therefore seem explanatorily irrelevant for metaphysically necessary fundamental moral facts such as UTILITARIANISM.

The second reason optimism about finding a third factor is misplaced, is that most non-naturalists explicitly reject the idea that fundamental moral facts can, even in part, be explained by an appeal to natural facts. Whatever the

[70] Enoch (2011, 146).
[71] This should be so across different views of the structure of moral explanation (cf. Fogal and Risberg 2020, 172–74). I am assuming here that 'utility' is a natural property.
[72] Berker (2019a). This is a simplified gloss of Berker's view.
[73] Fogal and Risberg (2020) provide further reasons why such a view is a bad fit for non-naturalists.
[74] Exactly how principles help explain particular moral facts is itself subject to controversy (Berker 2019b; Enoch 2019; Fogal and Risberg 2020).

fundamental moral facts are, non-naturalists will not seek to explain them, even in part, in terms of a further set of natural facts. This is particularly true given the way I defined non-naturalism in §1.4.

A non-naturalist picture of moral explanation therefore seems to rule out the possibility that there could be, even in principle, any third factor—in the form of a set of natural facts—that could perform the dual duty of explaining moral beliefs as well as fundamental moral facts.

A third reason the non-naturalist's optimism is misplaced, can be seen by considering the possibility of letting go of the assumption that the third factor must be some set of natural facts. Doing so would not help the non-naturalist, as fundamental moral facts are not explained, even in part, in terms of some further non-natural moral fact either. If UTILITARIANISM was explained, even in part, by some further moral fact, the principle wouldn't be a *fundamental* moral fact. So, for a non-naturalist, there is not much at all that could help explain fundamental moral principles while simultaneously contributing to the explanation of moral beliefs.

That the above is indeed a fair characterization of the commitments of a non-naturalist, and ones that a non-naturalist like Enoch would be hard-pressed to take issue with, can be seen by considering his own writings on the structure of moral explanations. In a recent paper, which is not concerned with the reliability challenge, Enoch sets out what a standard non-naturalist picture of moral explanation looks like. The upshot of that picture is that

> if utilitarianism is true, then the fact that you ought to maximize utility (or some such) is an ungrounded moral fact; and all other moral facts are partly grounded in it (and partly in the natural facts, facts about which line of action will have which consequences, and the like).[75]

Such a picture, according to Enoch, is

> a fairly natural picture even independently of Robust Realism. But there are reasons to think that for Robust Realism it's more than just natural—it's needed.[76]

If this is so, non-naturalists such as Enoch must hold that fundamental moral facts are not explained by further natural or moral facts. That might be fine as far as it goes—explanation must always end somewhere. But if that is the case, there can be no third factor, not even in principle, that plays the required dual

[75] Enoch (2019, 3).
[76] Enoch (2019, 4).

explanatory role needed of it to explain the MINIMAL CORRELATION. This, in and of itself, makes it impossible for the third-factor strategy to succeed, at least when wielded by a non-naturalist.

5.7 Conclusion

In this chapter, we have seen that the principles UNEXPLAINED RELIABILITY and NO COINCIDENCE can undergird an epistemological debunking argument in the form of a reliability challenge. The reliability challenge threatens to undermine moral beliefs unless their reliability can be explained. A central component of explaining the reliability of moral beliefs consists in explaining the correlation between one's own moral beliefs and the moral truths—what I have called the MINIMAL CORRELATION. I presented a taxonomy of explanatory models that can be used to explain that correlation—direct explanations, indirect explanations, accidental explanations, and full-blooded Platonism.

I then set out one strategy non-naturalists have employed to answer this challenge—the third-factor strategy. I showed that certain attempts at formulating third-factor explanations in the literature are misguided because they do not fit the unified, dual, explanatory structure required of a third-factor explanation. Having set out the structure such an explanation would need to have, I then considered whether the resulting third-factor strategy is question-begging. I claimed that while it is indeed question-begging, there are no clear grounds for holding it to be *objectionably* question-begging.

I then went on to argue that the third-factor strategy fails, not because it is question-begging, but because the explanatory model it relies on is incompatible with the structure of moral explanation that non-naturalists, by their own lights, should endorse. The third-factor strategy is incompatible with moral explanations because there is no suitable third factor that is capable of participating in the explanation of both our moral beliefs and the fundamental moral facts. This is not least because fundamental moral facts, for the non-naturalist, are explained neither by natural nor non-natural facts—they are ungrounded. It is therefore impossible for any fact to stand in the relevant explanatory relationship with them.

In the next chapter, we will look at a strategy that jettisons the third-factor strategy in favor of embracing an accidental correlation between our moral beliefs and the moral facts.

6 A Cosmic Coincidence? The Accidental Correlation Strategy

6.1 Introduction

The previous chapter argued that non-naturalists cannot appeal to the third-factor strategy when attempting to explain the reliability of moral beliefs. The non-naturalist is therefore back where she started—without an answer to the reliability challenge. Worse still, three out of the four available models for explaining the pertinent correlation between moral beliefs and moral facts—MINIMAL CORRELATION—have now been rejected. We have seen that there is no direct explanation to be had for non-naturalists; moral facts do not explain moral beliefs, nor vice versa. Neither is there some model of indirect explanation available—some third factor—that contributes to the explanation of both of the correlated variables. Lastly, an explanation of the correlation modeled on mathematical full-blooded Platonism has the unattractive feature of failing to allow for a uniquely true, non-theory relative, answer to the practical issue of how to act.

The only remaining explanatory model is taking the correlation between moral beliefs and moral facts to be accidental—the result of a lucky "cosmic coincidence." In this chapter, I explore whether a non-naturalist could appeal to this model when answering the reliability challenge. Let us call this strategy *the accidental correlation strategy*. Not surprisingly, the strategy faces the objection that taking the reliability of our moral beliefs to be the result of a cosmic coincidence is, on the face of it, *deeply* implausible.

If the non-naturalist is to salvage the accidental correlation strategy, they need to show that the coincidence that explains the reliability of moral beliefs is not the type of coincidence that can undermine them. Such a response demands a two-pronged defense. The first prong consists in making explicit what it is about coincidences that can undermine beliefs. Second, the proponent of the accidental correlation strategy must then show that moral beliefs, on their account, are not subject to *this* type of epistemic coincidence.[1] This chapter will evaluate this two-pronged strategy for defending the accidental correlation strategy.

I begin by making clear what the accidental correlation strategy entails (§6.2), before considering the central problem faced by it—a violation of No Coincidence. To evaluate the force of this objection, we will look at attempts to explain how coincidences can undermine beliefs. Standardly, epistemic coincidences are identified through the modal profile of beliefs. I consider two such modal conditions that are intended to block epistemically coincidental beliefs from either amounting to knowledge or from being justified.

First, I discuss the view that moral beliefs, if coincidentally reliable, would fail to be epistemically sensitive (§6.3). I then consider a descendant of the notion of epistemic sensitivity—epistemic safety—and the worry that the accidental correlation strategy will render moral beliefs epistemically unsafe (§6.4).

I argue that neither modal condition currently seems particularly threatening to the accidental correlation strategy. This, in turn, means that there are no obvious modal interpretations of 'epistemic coincidence,' where it can be furnished into an objection against the accidental correlation strategy. Insofar as epistemic coincidences should be given a modal interpretation, the prospects for the accidental correlation strategy are therefore far better than one might have expected. I end the chapter by tying up some loose ends in the discussions so far (§6.5), before summing up (§6.6).

6.2 The Accidental Correlation Strategy

The taxonomy of models for explaining correlations set out in §5.3.2 described an accidental correlation between two variables, A and B, as one where there is no unified explanation of the relationship between A and B. There is, in other words, no direct relationship between the two associated elements, and neither is there an indirect relationship whereby a third factor plays a dual explanatory role with respect to them both. The fact that the variables involved are associated at all is therefore a coincidence.

[1] An alternative strategy would be to identify a type of *epistemically unproblematic* coincidence and claim that the one involved in the reliability of moral beliefs is of this type.

Interestingly, David Enoch seems to endorse something close to the accidental correlation strategy in an attempt to tie up loose ends created by his own sketch of a third-factor explanation, discussed in the previous chapter.

> [An] obvious challenge is that [the third-factor] explanation itself seems to invoke what may be thought of as a miracle. For isn't it an amazing fluke that whatever evolution "aims" at happens to also be good? And isn't this itself something that calls for explanation, an explanation that the [non-naturalist] is not in a position to offer?[2]

In response to this lucky coincidence, Enoch argues that the non-naturalist is, on the one hand, allowed to take herself to be defeasibly justified in her belief that there are metaphysically necessary moral truths. If so, it is no miracle that truths about what is good or bad obtain—they do so in every possible world.[3]

On the other hand, Enoch argues that evolution could not easily have failed to induce the spread of traits that are adaptive and which secure survival or reproductive success for the organisms in its thrall.[4] It is therefore no surprise that evolutionary processes, at least in close possible worlds, have influenced agents like us to have pro-attitudes towards actions that tend to promote survival or reproductive success. If so, it is no miracle that organisms like us are disposed to hold beliefs such as that survival is good. The association between the two associated phenomena—moral beliefs and moral facts—therefore have fully separate, independent, and non-miraculous explanations.[5]

Enoch finds that, when combined, these two independent explanations provide a minimally—but sufficiently—acceptable explanation of the relevant coincidence, which he is willing to settle with.[6] This is therefore a case where the explanation of the reliability of our beliefs, at least to some extent, is explained by a cosmic coincidence.

As mentioned, Enoch employs this non-unified, accidental correlation in order to shore up loose ends produced by his third-factor explanation. He would therefore seem to employ both types of strategies, thereby incurring the individual costs of each. In the rest of this chapter, I will explore the prospects for pursuing the type of accidental correlation strategy that Enoch here hints

[2] Enoch (2011, 172).

[3] Enoch (2011, 172).

[4] Enoch (2011, 172–74).

[5] Clarke-Doane has discussed and defended this type of strategy in a number of publications (2015; 2016; 2017), although he has come to have some reservations about its prospects in recent writings (Clarke-Doane 2020). See also Olson (2019).

[6] Enoch (2011, 175).

at, but removed from the context of providing support for a third-factor explanation.

The strategy now under discussion—explaining the reliability of our moral beliefs by appeal to a fortunate, cosmic coincidence—has seemed to many like a non-starter. According to the accidental correlation strategy, all there is to say about the reliability of moral beliefs is that evolutionary selective pressures have influenced agents like us to form beliefs that, entirely coincidentally, significantly overlap in content with the independent moral truths.

Such a strategy opens up a non-naturalist to the objection that our moral beliefs are epistemically coincidental in a way that undermines them. More specifically, the strategy unabashedly runs afoul of NO COINCIDENCE.

Recall,

> NO COINCIDENCE
> For any theory, T, which concerns a domain, D, if T entails that the reliability of D-beliefs is merely coincidental, that strongly counts against T.

The prospects for the accidental correlation strategy depend on showing either that the principle is false or otherwise misguided, or showing that there is sufficient counterweighing evidence in favor of the accidental correlation strategy to outweigh its evidential force. In order to evaluate the prospects of the accidental correlation strategy, it is necessary to determine what undergirds the plausibility of NO COINCIDENCE.[7]

The underlying idea behind NO COINCIDENCE is that forming beliefs that you have reason to believe are reliable, but where this reliability can only be explained by appealing to a coincidence—and perhaps even a large, patterned coincidence—undermines them. Why should such a coincidence have the power to undermine the justification ones has for holding a belief? As a first pass, note that it is widely agreed among epistemologists that epistemic coincidences (or ‘epistemic luck’) can block a true belief from constituting knowledge, or defeat its justification.[8]

Consider Gettier cases, where a subject has a true, justified belief, but where some feature of the situation blocks the belief from amounting to knowledge.[9] Imagine a man who looks at a normally reliable clock that, unbeknownst to him, has stopped. As the time is in fact exactly what the stopped

[7] Some who, explicitly or implicitly, defend similar principles are Field (1989, 26), Huemer (2005, 123), Street (2006), Bedke (2009), and Enoch (2011, 160).
[8] Pritchard (2007). For a rare exception, see Hetherington (1998).
[9] Gettier (1963).

clock shows, the man forms a true, justified belief about what time it is.[10] However, because the belief is subject to a certain type of coincidence, the man's true belief arguably does not amount to knowledge.

Exactly what it is that goes wrong in Gettier cases is a contested issue, but it has been popular to attempt to capture the problematic aspect of epistemic coincidences in a modal framework. This approach has led to the formulation of various modal conditions on beliefs that are intended to rule out the problematic kinds of epistemic coincidence.[11] Such conditions have been proposed as the element that is missing from traditional accounts analyzing knowledge in terms of justified, true belief.

Debunkers have often, implicitly or explicitly, co-opted this type of modal conception of epistemic coincidences, and attempted to undermine the epistemic credentials of moral beliefs by demonstrating that those beliefs fail to satisfy some such modal condition. Many debunking arguments, such as Joyce's and Street's, could be taken to impugn the modal profile of moral beliefs indirectly. Such debunking arguments have tended to involve the claim that some, or all, important explanatory connections between moral beliefs and moral facts are missing. In turn, this is taken to show that the beliefs could easily have been wrong.

On this view of debunking arguments, they undermine moral beliefs by taking the lack of an explanatory connection to indicate a lack of modal covariation between moral beliefs and moral truths. The lack of modal covariation means that the moral beliefs fail to satisfy the relevant modal conditions, which in turn undermines their claim to constituting knowledge, or undermines their positive justificatory status. In the end, it is therefore the modal profile of a belief that does the heavy lifting in terms of undermining moral belief.[12]

Certain realists might seek to rebut such claims by showing that moral beliefs in fact *do* possess the explanatory connections that in turn indicate the appropriate modal covariation required for positive epistemic statuses such as knowledge or justification. A proponent of the accidental correlation strategy, however, grants from the outset that moral beliefs possess no such explanatory connections. If a lack of explanatory connections means that moral beliefs, because of their modal profile, fail to satisfy conditions for positive epistemic

[10] The case is found in Russell (2009, 91).

[11] See Faraci (2019) for a discussion (and critique) of what he calls 'modalist' accounts of epistemic coincidence.

[12] Korman & Locke (2020).

statuses such as knowledge or justification, this would spell trouble for the accidental correlation strategy.

Given this rough picture of what is problematic about coincidentally reliable beliefs, we can now see what could motivate a principle such as NO COINCIDENCE. If there are no explanatory connections between moral beliefs and the moral truths, and any overlap is merely a coincidence, this means that there is very likely *no robust counterfactual co-variation* between the content of our moral beliefs and the content of the moral truths.

The task for the proponent of the accidental correlation strategy is to show that our moral beliefs do not fail to satisfy any such requisite modal conditions. Below, we will look at two such modal conditions—epistemic sensitivity and safety—and consider whether they spell trouble for the accidental correlation strategy.

6.3 Debunking Arguments from Insensitivity

One challenge to moral beliefs on the basis of their modal profile concerns their purported lack of epistemic sensitivity.[13] In this section, we will evaluate whether a belief's lack of sensitivity is sufficient to undermine it, and whether debunkers have provided reason to think that moral beliefs fail to be sensitive.

What does it mean for a belief to be epistemically sensitive? In its simplest form, epistemic sensitivity requires that, for some subject S and proposition p, if p were false, S would not believe p.[14] To see how a debunking argument based on the sensitivity condition would go, consider the following invocation of the modal profile of moral beliefs from Michael Ruse.

> You would believe what you do about right and wrong, irrespective of whether or not a 'true' right and wrong existed! […] Given two worlds, identical except that one has an objective morality and the other does not, the humans therein would think and act in exactly the same ways.[15]

The claim expressed by Ruse is that the influences that shape our beliefs are responsive to what is adaptive, and, if there is no difference in what is adaptive

[13] Implicit or explicit appeals to sensitivity are found in Ruse (1986, 254), Joyce (2001, 163), Sinnott-Armstrong (2006, 43), and Braddock (2017).

[14] The sensitivity principle has its roots in the work of Alvin Goldman (1976) and Fred Dretske (1971), but was most famously defended by Nozick (1981).

[15] Ruse (1986, 254).

across the two worlds, agents in those worlds would have identical moral beliefs. When it comes to morality, the thought goes, truth is not required for moral attitudes to be adaptive. If this is so, it might seem that humans—in a counterfactual scenario—could have their moral attitudes shaped in just the way we have, despite the content of those beliefs being false.

If this claim is coherent and plausible, moral beliefs would fail to be sensitive to the moral facts in the sense that we would still have the same beliefs we actually do, even if they were false.[16] This means that, even if we assume that our beliefs are true, they could in a sense easily have been false because we are unable to discern whether they are true or false. Worries about the modal profile of moral beliefs therefore have teeth even if we assume the beliefs to be true. If lack of sensitivity blocks a belief from having a positive epistemic status, then such a debunking argument could uniformly undermine moral beliefs.

The sensitivity condition was originally intended as a necessary external, modal condition on knowledge. Most notably by Robert Nozick, who termed it the 'variation condition' and employed it as part of his so-called tracking account of knowledge.[17] If sensitivity is required for knowledge, then an argument such as Ruse's could show that we cannot have moral knowledge.

Rather than taking sensitivity to be a necessary condition for knowledge, one can take a belief's failure to satisfy the sensitivity condition to instead be a sufficient condition for undermining justification. Insofar as justification is required for knowledge, lack of sensitivity, through undermining justification, thereby blocks a belief from amounting to knowledge as well. If a belief's lack of sensitivity is sufficient for undermining it, an argument such as Ruse's could show that moral beliefs are uniformly undermined.

Debunking arguments could be developed on the basis of the sensitivity condition in either of these ways.[18] Note that the charge that moral beliefs fail to satisfy the sensitivity condition—that they are *insensitive*—is faced not only by proponents of the accidental correlation strategy. If we accept that sensitivity is required for knowledge, then any view that takes us to have moral knowledge would need to show that our moral beliefs are sensitive. The non-

[16] The widespread use of the term 'truth-tracking', which has been prominent in the debunking literature seems to, at least usually, have been shorthand for the feature captured by the sensitivity principle (cf. Kahane 2011).

[17] Nozick (1981, chap. 3).

[18] For a moral bunking argument explicitly taking the insensitivity be a sufficient condition for epistemic undermining, see Braddock (2017).

naturalist, and a proponent of the accidental correlation strategy in particular, might nonetheless be thought to face a steeper challenge on this score.

To evaluate the charge that moral beliefs are insensitive, it is necessary to consider the plausibility of the sensitivity condition itself, as well as whether debunking arguments provide us with reason to believe that moral beliefs are insensitive, even if they are true. I will go through these two issues in turn.

In its simplest form, the sensitivity principle is subject to straightforward counterexamples. Consider a case given by Nozick himself.[19] A grandmother is visited by her grandson, and she sees that he is well. One can stipulate further details of the case, such that it is overwhelmingly plausible that she *knows* that the grandson is well. However, had the grandson not been well, he would not have been present. When asked, the family would nonetheless have *told* the grandmother that he was well. Therefore, if the grandson were not well, the grandmother would still have believed that he was well. So even though she would clearly seem to know that he is well, her belief that this is so, is not sensitive. This would seem to show that sensitivity cannot be a necessary condition for knowledge.

Such cases can be handled by requiring that we evaluate the sensitivity of a belief only relative to a particular belief-forming method.[20] Insofar as the grandmother uses the same method she in fact uses (i.e., ocular inspection), she could not easily have believed her grandson to be well, while he in fact was ill. This refined version will then give us SENSITIVITY, introduced in Chapter 3. Recall,

> SENSITIVITY
> S's belief that p, formed via method M, is sensitive if and only if, if p were false, S would not believe that p via M.[21]

There are a number of objections to taking even this refined version of sensitivity to be a necessary condition on knowledge (or failing to satisfy it to be a sufficient condition for epistemic defeat). One is that it faces a dilemma of either giving rise to wide-ranging skepticism or denying a highly intuitive epistemological principle. Consider the belief that I have two hands. If it were false—say I only had one hand—but things otherwise remained unchanged, I would not believe that I had two hands. The belief that I have two hands is therefore sensitive, and I can know that I have two hands.

[19] Nozick (1981, 179).
[20] Cf. Becker (2012, 87–88).
[21] I borrow the particular formulation of sensitivity from Bogardus (2016, 639).

Consider now the belief that I am not a handless brain-in-a-vat. If *this* belief were false—and I in fact *were* an envatted brain without hands that were stimulated to experience the world just as I actually do—I would *still* believe that I had two hands. If I were a handless brain-in-a-vat, therefore, I would still believe that I was not. Hence, the latter belief—that I am not a handless brain-in-a-vat—is *not* sensitive. If sensitivity is required for knowledge (or insensitivity is sufficient for epistemic undermining), I cannot know that I am not a handless brain-in-a-vat. This point generalizes to all sufficiently skeptical scenarios and therefore blocks us from knowing a great many things.

One way around this issue would be to deny that knowledge is closed under known entailment (*closure*, for short). Closure can be understood as the claim that if you know that p, and know that p implies q, you know that q. If we deny closure, I could know that I have two hands without knowing that I am not a handless brain-in-a-vat. Nozick branded the denial of closure as a feature of his tracking account, as it helps avoid skepticism. Even so, most have considered the proposed cure worse than the disease. For instance, by denying closure one opens the way for what Keith DeRose has called "abominable conjunctions" such as "I know that I have two hands, but I don't know that I'm not a handless brain in a vat."[22]

The sensitivity condition faces further challenges as well. It has been argued to be incompatible with inductive knowledge, allowing for the possibility of knowing a conjunction without knowing its conjuncts individually, and failing to allow for certain types of higher-order beliefs to amount to knowledge.[23]

Given the sheer number of epistemological issues facing the sensitivity condition, most epistemologists have given up on championing it as a necessary condition on knowledge.[24] The claim that a belief failing to satisfy the sensitivity condition—i.e., being insensitive—is sufficient for epistemic defeat has its fair share of challenges as well. There seem to be cases where an agent *acknowledges* that their belief is insensitive, but where the belief nonetheless remains justified.[25]

If sensitivity is neither a necessary condition for knowledge, nor insensitivity a sufficient condition for epistemic defeat, it is not clear that moral beliefs would be undermined *even if* they were insensitive. This would make the

[22] DeRose (1995).

[23] Vogel (2012) contains a systematic treatment of many of these objections.

[24] For evaluations of the merits of the sensitivity principle in relation to knowledge and justification, see the articles in Becker and Black (2012).

[25] Vogel (2012, 130–31).

sensitivity condition incapable of explaining why it is implausible to hold the reliability of moral beliefs to be explained by a coincidence.

For the sake of argument, let us assume that failing to be sensitive *is* a sufficient condition for epistemic defeat. After all, it would seem that many intuitive cases of insensitive beliefs are undermined.[26] Even assuming this, a second task remains for the debunker. We would need some reason to think that moral beliefs are not sensitive if the the accidental correlation strategy is true. There are several reasons for thinking that establishing this is no easy task.

6.3.1 Metaphysically Necessary Moral Facts

The main difficulty for showing moral beliefs to be insensitive, even if true, comes from the fact that the statement of SENSITIVITY as a counterfactual tends to be interpreted in terms of a possible worlds heuristic like the following.

> SENSITIVITY*
> S's belief that p, formed via method M, is sensitive if and only if, in the nearest possible world(s) where p is false, S does not believe p via M.[27]

Such interpretations of the sensitivity principle have been known to be uninformative when applied to belief in metaphysically necessary truths. This is because when a counterfactual conditional has a necessarily false antecedent, the standard interpretation, following Lewis, is that it is vacuously true.[28]

If some moral facts, such as fundamental moral facts, are metaphysically necessary, then it is metaphysically impossible for them to not obtain. This means that the antecedent of SENSITIVITY—"if p were false"—is necessarily false, which in turn renders the conditional vacuously true. For that reason, any belief with a metaphysically necessary truth as its content (whether moral

[26] Though this could of course be explained, not by them being insensitive, but because they suffer some other epistemic failing which nonetheless often overlaps with being insensitive.
[27] Cf. Greco (2012, 194–95).
[28] Lewis (1973, chap. 1).

or otherwise) is therefore vacuously sensitive.[29] Sensitivity is therefore famously ill-equipped to handle belief in necessary truths.[30] At least, this is so when interpreted according to the standard Lewis–Stalnaker semantics for counterfactual conditionals.[31]

A debunker could make several replies to this. First, the debunker could deny that the fundamental moral truths are metaphysically necessary, or at least claim that it would be question-begging for the non-naturalist to assume them to be so. Second, the debunker could argue that whether or not some moral truths are metaphysically necessary, many moral truths are contingent, and our beliefs in contingent moral truths could still be insensitive. Third, a debunker could claim that we should adopt some non-standard semantics for counterfactuals which allows a belief in necessary truths to be insensitive. I will consider each of these replies in turn.

As a first reply, could a debunker simply deny the non-naturalist's claim that some moral truths are metaphysically necessary? This question should remind us of the debate over whether the third-factor strategy is objectionably question-begging, discussed in §5.5. The lesson learned there was that it is hard to rule out such an assumption merely because it is part of what is contested by the debunker. A debunker is of course free to introduce independent arguments for why moral facts are not metaphysically necessary. Launching such metaphysical arguments against the central tenets of non-naturalism is perfectly fine, but it would constitute a very different challenge than either the reliability challenge or a debunking argument from insensitivity.

Insofar as a debunker is pressing an epistemological challenge grounded in the non-naturalist's lack of an explanation of the reliability of moral beliefs, or their insensitivity, it seems that the assumption of metaphysical necessity on behalf of the non-naturalist is legitimate. However, as we will see in the next subsection, it might not matter all that much if the non-naturalist is granted the necessity of some moral facts.

[29] Some have recently taken issue with the consensus that metaphysical necessity is "the strictest real (non-epistemic, non-deontic) notion of necessity" (Clarke-Doane 2019, 266). If this assumption is rejected, and one introduces an even more inclusive notion of necessity, it is not necessarily the case that metaphysically necessary beliefs are vacuously sensitive.

[30] This is not surprising. Nozick included a further necessary condition on knowledge—*adherence*—to handle belief in necessary truths. For discussion of the adherence condition in a debunking context, see Risberg and Tersman (2019).

[31] Stalnaker (1968), Lewis (1973).

6.3.2 Contingent Moral Facts

The second reply from the debunker starts from the observation that whether or not some moral facts are metaphysically necessary, some are surely contingent. Beliefs in contingent moral facts could still be insensitive. Take the fact that it was wrong of John to set a stray cat on fire last Thursday. Call this fact M. There are possible non-M worlds, such as a world where John did not set any cat on fire last Thursday. A belief in the contingent moral fact M is therefore not vacuously sensitive. But is it likely that beliefs in contingent moral facts are insensitive?

To answer this, it will be instructive to return to the model of moral explanation that was set out in §5.6. That model held that what ultimately explains non-fundamental moral facts—i.e., contingent moral facts such as the fact that some particular action is wrong—is some fundamental moral principle together with some set of natural facts.

This means that if some moral fact M obtains in one world, but not in another, this cannot be on account of the fundamental moral principles being different across these worlds, as these are metaphysically necessary. It must therefore be because the natural features are different across these worlds.

When evaluating the sensitivity of a belief in a contingent moral fact F, we go to the closest non-M world. Because of the structure of moral explanation (or supervenience, if you would like), there will be an accompanying difference in natural features across M and not-M worlds.[32] Humans are generally adept at tracking changes in natural features, such as the facts about whether or not John set a cat on fire last Thursday. We can therefore expect that our beliefs would in fact *track* such changes in natural features. Justin Clarke-Doane voices this argument succinctly.

> [F]or any atomic moral belief that A is M, had A not been M, then A would have been different in non-moral respects […] Moreover, had A been different in non-moral respects, then our moral beliefs would have reflected the difference.[33]

[32] Similarly, some have recently argued that, in light of the nature of moral supervenience, our beliefs in contingent moral facts are very unlikely to be insensitive. For such arguments, see Wielenberg (2010), Mogensen (2014, 115–17), and Clarke-Doane (2020, 106). As supervenience of the relevant kind would be a downstream explanatory consequence of the structure of moral explanations set out in §3.6, I take these arguments as parallel to the one I set out in the text (Fogal and Risberg 2020 cf.).

[33] Clarke-Doane (2020, 106). Footnotes removed. By "atomic moral belief" Clarke-Doane has in mind a belief that "A is M, where A names a particular person, action, or event and M ascribes a moral property" (2020, 106).

Clarke-Doane therefore concludes that beliefs about contingent moral truths are not likely to be insensitive. At the very least, we have been provided no reason to think that they are. This line of thinking, I believe, holds true for belief in most contingent moral truths. But perhaps not all.

Some debunkers have argued that certain beliefs in contingent moral facts could be insensitive because more or less hardwired cognitive mechanisms block us from appropriately revising our beliefs despite a variation in natural features. If there are moral judgments that we are sufficiently strongly disposed—or even hardwired—to make, then it could be that we would not revise them despite a relevant change in natural features.

Consider a perceptual analogy. The Müller-Lyer illusion makes two arrow-like lines appear to be of different lengths despite in fact being equal. In such cases, one's perceptual judgment about the lines' unequal length can remain despite one's propositional belief that the lines are equal. Such illusions are cases where an evolved faculty—our perceptual faculty—is prone to making systematically false (and therefore trivially insensitive) judgments, and where it fails to track easily verifiable natural features. If there are similar cognitive mechanisms for making moral judgments, it might very well be that the resulting judgments do not properly covary with changes in natural features.

In the moral case, a similar phenomenon could occur if certain belief-generating processes behind moral beliefs are deeply rooted in non-cognitive processes, which makes them far less likely to be responsive to new information about relevant natural features.[34] Whether or not there are any such processes underlying moral belief generation is an empirical question that at present lacks a definitive answer.

There are, however, suggestive proposals. Jonathan Haidt and colleagues claim to have found a general unwillingness to reevaluate the moral status of actions such as consensual sibling incest, despite making all manner of stipulations about changes in natural features, such as the action not leading to a reduction in well-being for either party.[35] The subjects report not being able to provide any justification for their judgment, but hold on to it nonetheless.

I want to be clear that I don't set out the above as a confirmed case of either insensitive moral belief or hardwired processes of moral belief-generation. Rather, I merely want to showcase that it is, at the very least, conceivable that

[34] Joyce (2001, 164–65) discusses a case that could possess such features.

[35] Björklund et al. (2000); Haidt (2001, 814). Note that Royzman et al. (2015) provide reasons to doubt whether the subjects in the study in question were *really* convinced that the relevant changes in natural features obtained (e.g. that the incest was not harmful).

moral belief in contingent moral facts could be insensitive. The argument that even the *possibility* of such insensitivity is blocked by considering the structure of moral explanation, supervenience, or the tracking of natural features therefore fails. However, for any particular claim of insensitivity of the above type, there would need to be evidence that the beliefs are cognitively impenetrable, or at least highly cognitively resistant.

If some moral beliefs do tend to fail to covary with relevant variation in natural features, that singles out a group of beliefs in contingent moral facts that are candidates for being insensitive. For the sake of argument, let us assume that there is some such restricted set of insensitive moral beliefs. Richard Joyce, also considering the example of incest, thinks even such a restricted conclusion is sufficient for his skeptical purposes.

> One might object that this only shows that judgments concerning incest are unjustified and that this says nothing about judgments concerning stealing, slavery, etc. But the point of the evolutionary hypothesis is to emphasize that moral judgments come from a naturally selected *faculty*: to show that that faculty sometimes systematically generates unjustified judgments is to show that the faculty is unreliable *simpliciter*.[36]

This, however, is surely wrong. Consider again the analogy with visual illusions. Humans have a perceptual faculty that has been selected for. It generates systematically insensitive judgments, as in the case of many perceptual illusions. The beliefs in such illusions, furthermore, are not always responsive to additional information which falsifies them. That does not suffice to show that *all* our perceptual judgments are unreliable or undermined. If one can explain why a certain class of beliefs, whether moral or perceptual, fall prey to such mechanisms, then it would seem that those can be epistemically quarantined so as to not infect the whole class of beliefs with unreliability.

Moral beliefs, therefore, fall in between Clarke-Doane's sanguine optimism and Joyce's pessimism about the sensitivity of beliefs in contingent moral facts. Different moral beliefs will fit better within either of these molds. Many of our moral beliefs will track the non-moral features and co-vary appropriately in a way that renders them sensitive. Insofar as having certain pockets of systemically insensitive beliefs is not sufficient to undermine the whole domain of beliefs of which they are a part, this seems insufficient to ground a global moral debunking argument.[37]

[36] Joyce (2001, 165).

[37] It might be grounds for a local debunking argument, however.

6.3.3 Non-Standard Semantics for Counterfactual Conditionals

The above discussion has relied on a standard semantics for counterfactual conditionals. A debunker could claim that we should abandon such a semantics, where counterfactuals are only evaluated with respect to metaphysically possible worlds. Instead, they could argue, we should opt for some non-standard semantics which also includes metaphysically *impossible* worlds in the evaluation.[38] Street's discussion of moral possibility indicates some sympathy for this move.

> [A]s a purely conceptual matter, these independent normative truths might be anything. In other words, for all our bare normative concepts tell us, survival might be bad, our children's lives might be worthless, and the fact that someone has helped us might be a reason to hurt that person in return. Of course we think these claims are false—perhaps even necessarily false—but the point is that if they are false, it's not our bare normative concepts that tell us so.[39]

On the relevant type of semantics, there will be metaphysically impossible worlds that are relevant for the evaluation of counterfactual claims. The antecedent of SENSITIVITY is therefore no longer necessarily false for beliefs about fundamental moral truths. Our moral beliefs about fundamental moral facts could therefore turn out to be insensitive. This is the third possible reply on behalf of the debunker to the apparent vacuous sensitivity of moral beliefs.

The move to include metaphysically impossible worlds in the evaluation of counterfactuals raises several complicated issues. First, we would have to determine what our beliefs, moral or otherwise, would be in a metaphysically impossible world. More specifically, we would need to determine that our belief in a fundamental moral truth remains in the closest impossible world where that belief is false. I will assume that we can, at least for the most part, make sense of such evaluations, although we'll see below that this is not always the case.

Broadening the range of worlds that are relevant to the evaluation of sensitivity counterfactuals gives rise to a second issue as well. Given that we are now allowed to evaluate the sensitivity counterfactual with respect to impossible worlds, are there other beliefs that become insensitive? On a standard semantics, *every* belief in a necessary truth is vacuously sensitive. On a non-standard semantics, that is no longer the case. If we were to find that beliefs

[38] For such approaches to counterfactual conditionals, see Nolan (2013) and Brogaard and Salerno (2013).
[39] Street (2008b, 208).

that we would not want to render insensitive, are indeed rendered insensitive on such a non-standard semantics, that can be a reason to reject it. This worry, then, is one about generalization.

Clarke-Doane has posed such a challenge with respect to beliefs in certain types of metaphysically necessary truths, including beliefs about mathematics and metaphysics. Consider first mathematical truths. Is the belief that 1+1=2 sensitive? To answer that question, we would need to determine whether we would still believe that 1+1=2 in a metaphysically impossible world where it was false. It is difficult to be sure if this question is intelligible, but to the extent it is, many have sided with Hartry Field in holding that "we would have had exactly the same mathematical […] beliefs, even if the mathematical […] facts were different."[40] This would make mathematical beliefs insensitive. Pursuing the debate would lead us too far astray here, so I will merely register it as a worry.[41]

Consider now metaphysical beliefs. Take the belief that atoms arranged in a chair-wise fashion constitute a chair. Assume that this is true, and necessarily so, and that it is not a conceptual truth.[42] We can then ask whether the belief in this metaphysical truth about composition is sensitive. Presumably, it is not, since in the closest impossible world where it is false, we would still believe it. And so on for other beliefs in highly abstract necessary truths that are not conceptual. This would render such beliefs insensitive.

Perhaps the debunker is happy to throw out abstract metaphysics along with morality. This incurs a dilemma, however.[43] Imagine a debunker who has accepted the insensitivity of metaphysical claims concerning composition. The debunker nonetheless believes that he is sitting on a chair. Either the insensitivity, and consequent undermining, of our beliefs about composition in turn undermine judgments involving ordinary composite objects, or it does not. If it does, then this alternative semantics makes everyday contingent beliefs—like the debunker's belief that he is sitting on a chair—insensitive. In this case, both moral and everyday beliefs would be rendered insensitive, but this would be an unattractive result for a debunker.

On the other hand, assume the debunker argues that the insensitivity, and consequent undermining, of a belief in abstract metaphysical composition

[40] Field (2005, 85). Field made the same claim for logical beliefs as well. For reasons discussed below, this might be more controversial.

[41] See Clarke-Doane (2020, 138–42) for a sympathetic defense of Fields claim. See Joyce (2016b, chap. 7) for dissent.

[42] If you think this is a conceptual truth, you can substitute for it some other highly abstract metaphysical principle which is plausibly not a conceptual truth.

[43] Cf. Clarke Doane (2016, 27).

principles *does not* undermine everyday beliefs about chairs. The debunker is then claiming that the insensitivity of beliefs in abstract metaphysical truths fails to render everyday beliefs that are partly explained by these truths insensitive. There now arises a question of why the same is not true of moral beliefs.

If the same is true of moral beliefs, a debunker could then show that beliefs in abstract moral principles, such as pain being bad, are insensitive, without undermining everyday moral beliefs of the type that it was morally wrong of John to set a stray cat on fire last Thursday. This would seem like an underwhelming result for a debunker.

The debunker therefore seems forced to find some relevant difference between the pertinent type of moral and metaphysical truths. But it is not clear what that difference would consist in. In the absence of some such difference, the debunker seems forced to choose between debunking too much—including mundane everyday beliefs—and debunking too little and leaving mundane everyday moral beliefs unscathed.

Lastly, in addition to such worries about generalization, I believe there is also a worry about self-defeat lurking in the shadows of this third reply. If some semantics for counterfactuals involving impossible worlds is sufficient to establish the insensitivity of fundamental moral beliefs, we need to apply the same framework when evaluating the sensitivity of our epistemic beliefs.

Consider the purported fact that a lack of sensitivity is sufficient to undermine justification. If true, it has a good claim to being an a priori, metaphysically necessary truth. Furthermore, it is plausibly not a conceptual truth.[44] Would we still hold this belief in the closest impossible world where it was false? To the extent that the question is intelligible, I cannot see any reason why we should think not. If our epistemic beliefs, such as beliefs about insensitivity, were to themselves be rendered insensitive, and consequently undermined, the debunker would face self-defeat.

The above line of reasoning puts pressure on those who defend debunking arguments utilizing the sensitivity principle coupled with a non-standard semantics for counterfactual conditionals to find some relevant difference in these cases. If they cannot, our epistemic beliefs would seem to be rendered

[44] If you do believe that at least some of the fundamental epistemic truths are conceptual truths, you could argue that a world where the fundamental epistemic facts are different, or do not obtain at all, is not even conceptually possible. If one accepts this claim about epistemology, there is plausibly pressure to accept the same claim about morality as well (cf. Kyriacou 2018). The moral and the epistemic domains would therefore be vulnerable to debunking arguments from insensitivity to the same extent.

insensitive as well, and as a result, the debunking argument from insensitivity would be self-defeating.

The adoption of a non-standard semantics for counterfactuals for the purpose of showing that moral beliefs are insensitive risks generalizing. It risks generalizing to domains where debunkers would presumably not want it to, and it furthermore risks generalizing in such a way that the argument becomes self-defeating. At the very least, adopting a non-standard semantics would seem to shift several explanatory burdens onto the debunker.

Summing up, debunking arguments from insensitivity run into four main problems. First, it does not seem that sensitivity is a necessary condition for knowledge, nor that insensitivity is sufficient for undermining justification. Second, belief in necessary moral truths is vacuously sensitive, and the non-naturalist's moral beliefs concerning fundamental moral facts will therefore be vacuously sensitive. Third, most—but perhaps not all—beliefs in contingent moral facts will likely be sensitive, if true, because we are able to track the relevant changes in natural features that necessarily go along with a change in moral facts across worlds.

Fourth, the sensitivity principle, if combined with a non-standard semantics for counterfactual conditionals, raises serious problems for debunkers. At best, it would shift a large explanatory burden onto the debunker, who will need to defend various views about the epistemology of mathematics as well as the nature of epistemological principles. At worst, it will cause their argument to overgeneralize and render other domains of belief insensitive as well, including mathematical beliefs, metaphysical beliefs, and ordinary everyday beliefs. It might even generalize to epistemic beliefs, which would make a debunking argument from insensitivity self-defeating.

The number of serious issues facing debunking arguments from insensitivity does not inspire confidence in their success.[45] While taking insensitivity to be a sufficient condition for epistemic undermining makes good sense of how many debunking arguments have been formulated, it does not allow them to succeed. A debunker would therefore be well advised to search for a different principle to underwrite their argument. Let us therefore move on to a different candidate modal principle.

[45] One might think that the sensitivity condition could be modified to circumvent these issues. This has been attempted by Braddock (2017), who has formulated a debunking argument utilizing an insensitivity principle that he claims can avoid the issues discussed so far. However, when targeting someone who subscribes to the picture of moral explanation set out in §5.6, Braddock's argument straightforwardly fails for the reasons set out in §6.3.2.

6.4 Debunking Arguments from Lack of Safety

The epistemological objections to the sensitivity principle have made it somewhat unpopular among epistemologists. In its wake, a closely related condition, epistemic safety, has surged in popularity. While sensitivity, in its simplest form, required of us that we would not believe that p, if p were false, safety is its contrapositive. It requires, in its simplest form, that we would believe that p, only if p were true.[46] In this section, we will consider whether safety is plausibly a necessary condition on knowledge, or if failing to satisfy the safety condition is a sufficient condition for epistemic defeat. We will also consider whether debunkers have provided any reason to think that moral beliefs fail to be epistemically safe.

As with sensitivity, there are a multitude of formulations of the safety condition in the literature.[47] Consider first a standard version of safety spelled out in a possible worlds framework.

> SAFETY
> S's belief that p, formed via method M, is safe if and only if, in close possible worlds where S believes that p via M, p is true.

Unlike sensitivity, safety is restricted to *close* possible worlds. This reflects the idea that the belief, if safe, could not *easily* have been false. A belief that p can therefore be safe even if one falsely believes that p in a far-off world. As with sensitivity, there is also a need to relativize the condition to a method.[48]

The notion of epistemic safety was first developed by Ernest Sosa as an alternative to, and improvement upon, sensitivity.[49] Like the latter condition, safety was initially intended to be a necessary condition for knowledge (and jointly sufficient together with true belief).[50] Since then, safety has come to be the dominant candidate for being a necessary external, modal condition on knowledge.[51]

[46] A conditional (A→B) and its contrapositive (not-A→not-B) tend to be logically equivalent, but this is not the case for counterfactual conditionals (Lewis 1973, 35). Safety is therefore not equivalent to sensitivity, despite being its contrapositive (Sosa 1999, 149–50 fn. 1).

[47] Rabinowitz (n.d.).

[48] Cf. Greco (2012, 195–96).

[49] Sosa (1999).

[50] Sosa has since abandoned a straightforward safety condition on knowledge.

[51] Proponents who have, at least at some point, endorsed it in their accounts of knowledge include Sosa (1999), Williamson (2000), Pritchard (2005; 2009).

Instead of taking safety to be a necessary condition for knowledge, it is possible to take a belief's lack of safety—the belief being *unsafe*—to be a sufficient condition for undermining its justification. Insofar as justification is required for knowledge, a lack of safety, through undermining justification, thereby blocks an unsafe belief from amounting to knowledge as well. Debunking arguments could be developed on the basis of the safety condition in either of these ways.

It is not as easy to find extant debunking arguments that explicitly rely on the safety principle. Even so, several authors do seem to present arguments that rely on its gist, by being motivated by the worry that our moral beliefs could easily have been false.[52] Such arguments often highlight the connection between the evolutionary genealogy of human moral beliefs and the fact that, if that genealogy had been different, we could have had different moral beliefs. Consider for instance the following from E. O. Wilson and Michael Ruse taken from a popular science magazine.

> Suppose that, instead of evolving from savannah-dwelling primates, we had evolved in a very different way. If, like the termites, we needed to dwell in darkness, eat each other's feces and cannibalise the dead, our epigenetic rules would be very different from what they are now. Our minds would be strongly prone to extol such acts as beautiful and moral. And we would find it morally disgusting to live in the open air, dispose of body waste and bury the dead. Termite ayatollahs would surely declare such things to be against the will of God. […] Ethics does not have the objective foundation our biology leads us to think it has.[53]

Joyce also frames his argument in terms of the possibility of diverging evolutionary histories at certain points.

> Were it not for a certain social ancestry affecting our biology, the argument goes, we wouldn't have concepts like obligation, virtue, property, desert, and fairness at all. If the analogy is reasonable, therefore, it would appear that once we become aware of this genealogy of morals we should (epistemically) […] cultivate agnosticism regarding all positive beliefs involving these concepts until we find some solid evidence either for or against them.[54]

Similarly, Rosenberg writes:

[52] Bogardus (2016, 645–46) catalogs a number moral debunking arguments which can be understood to be motivated to safety-adjacent worries. Beyond morality, there are debunking arguments against religion (Wilkins and Griffiths 2012) and virtue epistemology (Olin and Doris 2014) that can plausibly be interpreted as employing a safety condition (cf. Kallberg 2021).
[53]Wilson and Ruse (1985), quoted in Bogardus (2016, 645).
[54] Joyce (2006, 181), quoted in Bogardus (2016, 646).

> If the environment had been very different, another moral core would have been selected for […] But it wouldn't have been made right, correct, or true by its fitness in that environment.[55]

Such arguments require some reconstruction in order to fit into the mold of a debunking argument from lack of safety. The general line of thought in such arguments would seem to be that, in our evolutionary past, things might have taken a different turn, and as a result, because different selective pressures, we might have had different, and incompatible, beliefs. We might therefore worry that our moral beliefs could have been false, even if we assume them to be true.

On its own, this conclusion would seem to merely underline the fact that there are stance-independent moral facts, such that, if our current beliefs are true, and we could have had alternative beliefs that would be *incompatible* with our current beliefs, then these alternative beliefs would have been false.

In order to point to a lack of safety with respect to our current beliefs, it would be necessary to show that we could have *easily* had such different, and incompatible, beliefs using the method of belief formation we in fact employ. If it is possible to construct such scenarios, and if the safety condition is either a necessary condition for knowledge or a sufficient condition for epistemic defeat, then such arguments could show that our moral beliefs uniformly fail to amount to knowledge or that they are uniformly undermined.

For a debunking argument built around SAFETY to succeed, two things are required. First, a belief being unsafe needs to either block the belief from constituting knowledge or be sufficient for epistemic undermining. Second, it needs to provide us with some reason to think that our moral beliefs are unsafe. We will consider these two issues in turn.

Whether safety is necessary for knowledge, or whether a lack of safety is sufficient for epistemic defeat, is currently subject to lively debate. Against such claims, some argue that it is possible to know that p despite one's belief that p being unsafe.[56] This, in turn, tends to generate ever more refined versions of the safety principle, often resulting in it being embedded in a virtue epistemic account of knowledge that also has further conditions concerning various epistemic capacities, abilities, or skills.[57] This again results in more

[55] Rosenberg (2012, 113), quoted in Bogardus (2016, 646 fn. 28).

[56] Neta and Rohrbaugh (2004); Comesaña (2005); Kelp (2009); Bogardus and Maxen (2014); Bogardus (2016); Lutz (2020).

[57] E.g. Sosa (2009), Pritchard (2012).

and more refined counterexamples. Pursuing this debate to any serious extent here would take us too far astray.

For our purposes, we can note that it is contested whether a belief being unsafe is sufficient for blocking it from constituting knowledge, or is sufficient for epistemic defeat. Taking on such a contested principle could therefore be a cost in itself for developing a successful debunking argument. However, as most principles will be controversial to some extent, let us for argument's sake assume that the safety principle can be defended in either of these ways.

Let now us consider whether debunking arguments provide us with a reason to hold that moral beliefs would fail to satisfy SAFETY. Note first that the formulation of the safety condition shares many features with SENSITIVITY. One of these is the lack of sophistication when it comes to handling belief in metaphysically necessary truths. A belief in a necessary truth *could not fail* to satisfy SAFETY. Whenever one forms a belief that p in a close possible world, where p is a necessary truth, p is guaranteed to be true. A belief in a necessary truth is therefore vacuously safe.[58]

Second, SAFETY faces similar challenges with respect to belief in contingent moral facts as those we saw for SENSITIVITY. Take the contingent moral fact that it was wrong of John to set a stray cat on fire last Thursday. Call this fact M. For a belief that M to be unsafe, we would need to form the belief that M in a close possible non-M world. The fundamental moral facts—the wrongness of inflicting unnecessary pain, say—is vacuously true in all close possible worlds. This means that for the belief that M to be false, there needs to be a change in natural features in this close possible world. For instance, the closest possible non-M world could be one where John did not set any stray cat on fire last Thursday.

For reasons that were discussed at length in §6.3.2, most beliefs in contingent moral facts are likely to be safe simply because of the nature of moral explanation (or supervenience). We are unlikely to fail to track the changes in natural features that accompany the change in truth value for a given moral proposition.

A straightforward version of the safety condition is therefore not a promising candidate for being the principle operative in debunking arguments or for underwriting the plausibility of NO COINCIDENCE.

In recent years, the safety condition has been modified in order to be more informative in its handling of belief in necessary truths. To this end, it has

[58] Such considerations led some, such as Pritchard (2005), to restrict the application of safety to beliefs in contingent facts. Pritchard has since modified his view.

been modified to allow the evaluation of beliefs with a range of propositional contents, and not only the content of the belief held in the actual world.[59] Given such modifications, we can formulate a refined safety condition as follows.

> SAFETY*
> S's belief that p, formed via method M, is safe if and only if, in every close possible world where S believes that p (or some sufficiently similar proposition p*), via method M, p (or p*) is true.

Such a formulation of the safety condition makes it necessary to evaluate more than the modal robustness of p in order to determine if the belief is safe. It requires that one also consider whether we, by employing the same method, could have formed a different but sufficiently similar belief that is false in some close possible world. When formulating safety in this way, it becomes necessary to determine when the content of two propositions counts as 'similar'. This constitutes a problem all of its own, but I will set it aside here.

To see how this change in the safety principle would make a difference, consider a case where you make up your mind about what the 999th prime number is. By taking a wild guess, you form the belief that it is 7907. As it happens, your belief is true! According to SAFETY, your belief is safe since there is no close possible world where you could have falsely believed that 7907 is the 999th prime.

According to SAFETY*, your belief is *not* safe, because there is a close possible world where you would have formed a similar, but incompatible belief, using the same method you in fact used. Perhaps you instead formed the belief that the 999th prime number is 7901. You formed this belief by employing the very same method of belief formation as you did in the actual world—taking a wild guess. Hence your actual belief fails to satisfy SAFETY* because the method you used to form that belief would, in a close possible world, lead you to form a sufficiently similar, but false belief.

I am not aware of any sustained discussion that argues that moral beliefs, in particular, fail to satisfy SAFETY*.[60] The issue is, unfortunately, severely underexplored. On the one hand, this could be taken to signal that a non-naturalist need not worry. Until there is some debunking argument based on

[59] Cf. Dunaway (2017).

[60] Clarke-Doane (2020) briefly considers the issue and Kallberg (2021) considers the extent to which human beliefs *in general* are able to satisfy a principle like SAFETY*. We will return to their discussions below.

SAFETY* that is formulated, there has been no reason provided to fear for the safety of moral beliefs. Even so, I think there are reasons to question whether concerns about safety are likely to pose a threat to the non-naturalist. I will suggest four points that speak against the prospects of a global moral debunking argument based on a principle like SAFETY*.

First, debunkers themselves have argued that our beliefs are quite robustly formed, which could fix our moral beliefs in close possible worlds. Second, there is the problem of individuating belief-formation methods, which is essential to the formulation of SAFETY*. Third, there is the problem that a debunking argument from safety might generalize. I will consider these in turn.

For SAFETY* to show that our moral beliefs are unsafe, it needs to be the case that p (or a similar proposition p*) is false in a close possible world where we believe that p (or a similar proposition p*).

The first obstacle to moral beliefs being unsafe in either of these ways is that—as we saw in Part I—many debunkers hold that our moral beliefs are influenced by evolutionary selection pressures to such an extent that these beliefs would be generated quite robustly. As Clarke-Doane has argued, debunking arguments' use of this premise can be turned around to create an obstacle to establishing a lack of safety.[61]

In arguing for why our beliefs are insensitive, for instance, proponents of evolutionary debunking arguments often claim that we were highly likely to come to possess at least some of the moral beliefs we in fact do, simply in virtue of them being evolutionarily beneficial.[62] If selection pressures that were likely to arise for agents like us can be used to establish the modal stability of moral beliefs, there might not be any close possible worlds where we would have formed different, and incompatible, beliefs using the methods we in fact do.

Insofar as the debunking argument in question itself relies on this claim, it allows the realist to ask if it is true, after all, that the content of our moral beliefs could *easily* have been different, and incompatible, with our actual beliefs. To the extent that our moral beliefs have been significantly influenced by evolutionary forces, and to the extent that these forces are modally robust, debunkers' own premises might provide some reason against thinking that our moral beliefs are unsafe.

In response to this, it is possible to imagine that evolutionary pressures would have been significantly different, even in close possible worlds, and

[61] Clarke-Doane (2016, 29; 2020, 109–10).

[62] This is central to e.g., Street's (2006) argument.

therefore allowed moral belief to diverge significantly from our actual beliefs. This kind of claim, however, would need to be substantiated.[63]

A second obstacle can be seen by recalling the argument from Wilson and Ruse above. As they claim, it is surely true that our cognitive states would be very different if our evolutionary ancestry had taken a significantly different turn and selection pressures were significantly altered.

In response, one might both ask whether such turns could have *easily* taken place, and if so, whether they would have resulted in us using the *same* method for forming our beliefs as we currently do. These factors lead to a tension that arguments from safety will need to navigate. On the one hand, they will need to claim that the selection pressures impacting our moral beliefs could have *easily* taken a *sufficiently different trajectory*. On the other hand, they need to argue that even after having gone down this significantly different trajectory, we would be employing identical belief-forming methods when forming moral beliefs as we currently do.

This brings up a notorious problem in epistemology—the problem of individuating methods—called the *generality problem*. Individuating methods is an immense problem for both sensitivity and safety theorists.[64] In fairness, the generality problem is a problem for almost everyone. Even so, if the question of whether moral beliefs are safe or not depends on the issue of how to individuate belief-forming methods, it might afford non-naturalists an easy way to a dialectical stalemate.

Under significantly different environments and selection pressures, it might be possible to argue that we are not forming our beliefs using the same methods we in fact use. At an extreme, consider Wilson and Ruse's scenario where, "like the termites, we needed to dwell in darkness, eat each other's feces and cannibalize the dead." It is at least conceivable that under such circumstances, our cognitive machinery would have developed along sufficiently dissimilar lines that we would no longer be forming beliefs in the same way we do now. To take just one example, under the specified circumstances, the human perceptual system, in response to the lack of sunlight, would likely be very different. Perhaps we would employ echolocation or some similar method of orienting ourselves.

Such large-scale changes in our perceptual system could radically reconfigure how humans form beliefs. If something similar could be said about the processes underlying moral belief generation in such alternative scenarios,

[63] For some discussion the possibility of alternative evolutionary pressures, see Mogensen (2014, 83–87).

[64] See Becker (2012) for discussion.

then that would block such scenarios from rendering moral beliefs unsafe. In addition, it will be necessary to show that the beliefs that are formed by an identical method, after having taken this significantly different trajectory, have sufficiently similar content.

These three tasks—establishing the possibility of alternative trajectories we easily could have taken, individuating methods, and establishing similarity in content—require substantial work. When taken together, they also pull in different directions. The more radical the alternative trajectory, the less likely the sameness of method (and perhaps also similarity of content). At the very least, establishing that moral beliefs are unsafe would require a substantial effort on behalf of a debunker.

Lastly, and perhaps most importantly, there is the issue of whether a debunking argument from SAFETY*, if successful, would debunk *too much*. As with sensitivity, any formulation of a debunking argument employing the safety condition must be careful so as to not result in an argument that targets too wide a swath of our beliefs.

How would SAFETY* fare in this regard? The few extant discussions of debunking arguments employing some principle like SAFETY* hold that, if successful, they would generalize beyond the moral domain. Clarke-Doane has argued that if moral beliefs are unsafe, so are mathematical beliefs as conceived by a mathematical Platonist, barring the adoption of full-blooded Platonism.[65] Kallberg has argued, even more radically, that debunking arguments from SAFETY*, if successful, would risk generalizing to *all human beliefs*.[66]

Here, we should recall the recurring worries about self-defeat that has attached itself to a number of the debunking arguments we have considered so far. To avoid self-defeat when employing the SAFETY* condition, a debunker would need to show that the reasoning behind impugning moral beliefs with a lack of safety does not equally impugn epistemic beliefs—including a belief in the safety condition itself. I will develop some concerns about the prospects for this task in the next two chapters.

The obstacles set out for debunking arguments from SAFETY* above are not insurmountable, at least not in principle. It could be that a debunking argument based on the safety condition could navigate them successfully. What we can safely note is that extant debunking arguments have not provided sufficient reason to think that our moral beliefs are unsafe. That is to say, our

[65] Clarke-Doane (2020, 152). For discussion of full-blooded Platonism, see §5.3.2.
[66] Kallberg (2021, 243).

moral beliefs could very well be unsafe, but debunkers have not provided sufficient evidence that this is the case.

Interestingly, a fully developed argument from lack of safety would, of necessity, need to take account of findings from evolutionary theory. To determine whether we could have *easily* formed moral beliefs that are similar, but incompatible, with our actual beliefs given identical methods of belief formation, would require close attention to actual and possible environmental and selective factors that have, or could have, influenced moral belief formation in humans.

Let us now take stock. We have seen that a debunking argument from a lack of safety is more promising than an argument from insensitivity. Despite this, we do not have a substantial, fleshed-out debunking argument based on SAFETY* that is intended to target the moral domain specifically, without afflicting other domains. This is sure to change in the years to come.

We have seen that, at present, the arguments that do consider whether moral beliefs are vulnerable to the charge of being unsafe are either underdeveloped or, seemingly, overly successful. If the cost of adopting the safety condition as a sufficient condition for epistemic defeat is that our beliefs about domains such as mathematics, or even most or all beliefs, become unsafe, that would strongly count against the safety condition.

Another thing to note about arguments from lack of safety is that they are a far cry from most debunking arguments that have previously been discussed. This means that if *this* is the form debunking arguments should take, then it is not the case that previous debunking arguments have been much of a precursor to it. Most of the mentions of alternative evolutionary trajectories are arguably tangential to the core of arguments like that of Joyce and Street. This means that while safety is perhaps the only plausible candidate for a principle that could be employed by a debunking argument considered so far, it is not a very close interpretation of previously formulated arguments.

Let us now return to the evaluation of the accidental correlation strategy.

6.5 Tying Up Loose Ends: Modal Conditions and Epistemic Coincidence

Return now to the accidental correlation strategy. This strategy holds that the reliability of our moral beliefs is entirely coincidental, in the sense that there is no unified explanation of the MINIMAL CORRELATION. Instead, it suggests that there is only a coincidental, non-unified explanation, which explains each associated variable separately. This violates NO COINCIDENCE, which gives voice to the intuition that explaining the reliability of some set of beliefs by appealing to a coincidence can only be a last resort.

By now, it should be clear that whatever it is that is objectionable about the accidental correlation strategy, and violating NO COINCIDENCE, it is rather hard to pinpoint. We have considered the attempt to explain it through the use of modal conditions—sensitivity and safety—that could serve to explain why violating NO COINCIDENCE would be sufficient to undermine moral beliefs. We have seen that sensitivity does not seem capable of playing this role because it is neither a sufficient condition for epistemic defeat nor does debunking arguments provide a reason to think that our moral beliefs are insensitive to a significant degree.

As for the safety condition, it might be a relatively plausible candidate for either being a necessary condition for knowledge, or a sufficient condition for epistemic defeat. But at present, there is little work that attempts to show that moral beliefs are unsafe. Furthermore, to the extent that our moral beliefs are unsafe, it seems likely that so are many other forms of beliefs. Because such generalization is likely to be unwelcome, and in turn reflects badly on the plausibility of the safety condition, there is no clear objection to the accidental correlation strategy from the safety condition at present. For safety to play this role, there would need to be evidence of our moral beliefs being unsafe without such evidence also generalizing to other unwelcome domains.

Insofar as neither of these modal conditions, at least at present, seem to impugn our moral beliefs, they cannot capture what it is about the accidental correlation strategy that is epistemically problematic. Given this, it is not clear what the charge against the accidental correlation strategy is supposed to be. Sure, it seems *prima facie* unintuitive, preposterous even, and one has the sneaking suspicion that *something* is wrong with it, epistemically speaking. But that is an unsatisfactory answer to why we cannot accept coincidentally reliable beliefs.

If this is so, it seems as if the non-naturalist has a way of countering the objection that the accidental correlation strategy is deeply implausible because it violates NO COINCIDENCE. That counter consists in claiming that while it is true that the reliability of our moral beliefs *is* coincidental, the coincidence in question is not epistemically problematic, as we have not seen any clear reason to doubt the sensitivity or safety of moral beliefs.

NO COINCIDENCE, on this view, should be understood as restricted to epistemically problematic coincidences. Why, after all, would it be implausible to have beliefs for which we have defeasible justification, and which, if true, would not seem to engender any apparent worries about their modal profile?

This situation leads to an interesting dialectical upshot. Epistemological challenges, including debunking arguments, have been thought to force non-naturalists to adopt theoretically costly positions, such as postulating faculties capable of rational intuition or adopting some custom-tailored epistemological view about quasi-perception. If the picture I have set out is correct, this situation is now turned on its head. The debunker finds herself in the unenviable situation of needing to settle debates over the correct view of the analysis of knowledge and justification in order to get on with her argument.

The non-naturalist, on the other hand, merely needs to poke holes in any conditions proposed by the debunker. Such poking of holes, of course, is a far easier enterprise than constructing the conditions themselves. The non-naturalist can then leave the constructive part to the epistemologists.

The person who has most fully developed the general strategy explored in this chapter is Clarke-Doane, who has argued that once any worries about sensitivity and safety have been fended off, there does not seem to be any further epistemological worries that spring from the reliability challenge for the non-naturalist.[67] More specifically, he argues that there *could not be* any argument that both leave the sensitivity and safety of our moral beliefs unchallenged, while also succeeding in undermining them.[68] The general picture endorsed by the accidental correlation strategist, after all, is one where our moral beliefs "were (all but) bound to be true."[69] Whatever epistemological theory one prefers, how could one claim to find an epistemic fault with such beliefs?

This might seem to situate the non-naturalist in the comfortable position of only having to defend the sensitivity and safety of moral beliefs, which, as we have seen, does not seem like an insurmountable challenge. Of course, the debunker need not give up. On the one hand, a debunker could try to modify the arguments relying on sensitivity, or develop arguments grounded in epistemic safety. On the other, they could attempt to appeal to a different, non-modal epistemological principle that could explain how genealogical considerations could uniformly undermine moral beliefs.

In the next chapter, we will look at an instance of the latter strategy. I will evaluate epistemological principles that are capable of explaining how evolutionary considerations can undermine our moral beliefs, despite not directly challenging their modal profile.

[67] Clarke-Doane (2016, sec. 2.4).

[68] Clarke-Doane (2016, 25 n. 10) takes this type of undermining to be undercutting defeat.

[69] Clarke-Doane (2015, 97).

6.6 Conclusion

In this chapter, we have explored the fourth and last explanatory model for explaining the MINIMAL CORRELATION. That model holds the correlation to be the result of a cosmic coincidence. The accidental correlation strategy makes no recourse to a third factor or any other explanatory relation between moral facts and moral beliefs. In turn, this opens the strategy up to the objection that a belief that is only accidentally true suffers some epistemic defect.

We have explored two modal conditions—sensitivity and safety—that are intended to capture this sense of epistemic coincidence and show that it undermines the afflicted beliefs, or blocks them from constituting knowledge. We saw that arguments relying on the sensitivity condition face two significant obstacles. First, it is not clear that the principle is independently plausible, neither as a necessary condition for knowledge, nor as a sufficient condition for epistemic defeat. Furthermore, even if it were independently plausible, a debunker would need to provide some reason to think that moral beliefs fail to satisfy the sensitivity condition. This seems challenging to do. The prospects for accomplishing both these tasks simultaneously—defending a plausible version of the condition, as well as showing that moral beliefs fail to satisfy them—seem difficult enough that it poses no immediate threat to the non-naturalist.

As for the safety condition, it might be a plausible candidate for either being a necessary condition for knowledge, or a sufficient condition for epistemic defeat. But at present, there is little work that attempts to show that moral beliefs are unsafe. Furthermore, to the extent that our moral beliefs are unsafe, so would many other beliefs seem to be. Because such generalization is likely to be unwelcome, and in turn reflects badly on the plausibility of the safety condition, there is, at least at present, no clear objection to the accidental correlation strategy from the safety condition.

Given this, it seems that the non-naturalist has a theoretically lightweight response to the reliability challenge that can be formulated without concocting some implausible moral epistemology or by otherwise making costly theoretical commitments. This means that the epistemological principles that are commonly taken to explain how our moral beliefs are undermined by failing to satisfy some modal condition are not successful.

7 Harman's Return: The Explanationist Challenge

7.1 Introduction

A common theme in previous chapters has been that genealogical explanations of moral beliefs put pressure on the idea that there is some explanatory connection between moral facts and moral beliefs. In Part I, we saw arguments to the effect that there is either no indispensable explanatory connection or even no such connection at all. In Chapter 5, we looked at and rejected views to the effect that there is such an explanatory connection, but that it is indirect, mediated by a third factor. In Chapter 6, we then explored arguments to the effect that while there is no such *explanatory* connection, that fact need not undermine moral beliefs as long as there is an appropriate *modal* connection.

This chapter will explore an attempt at pushing back against the latter "modalist" view that a debunking argument could only undermine moral belief through challenging the modal status of those beliefs. In particular, we will consider an argument that holds that acknowledging the lack of explanatory connections is a sufficient condition for undermining the justification of a belief.

I start by introducing the explanatory constraint introduced by Harman and endorsed by Joyce (§7.2). I then present a recent formulation of an explanationist debunking argument due to Daniel Korman and Dustin Locke that utilizes an explanatory constraint on rational belief (§7.3). I argue that such constraints face two related problems. First, they are likely to generalize to domains such as mathematics, metaphysics, modality, and logic, and perhaps even empirical domains such as facts about the future (§7.4). I then show that such constraints seem, *prima facie,* to target belief in epistemic facts as well. If so, the argument risks being self-defeating since it relies on epistemic premises (§7.5). I discuss how such debunking arguments can avoid self-defeat and show that conditional and non-conditional debunking arguments have very different resources available for doing so. I end by summarizing and concluding (§7.6).

7.2 Explanationism, Again

Debunking arguments have tended to argue that some, or even all, important explanatory connections between moral beliefs and moral facts are missing. In turn, this is taken to show that the beliefs could easily have been wrong. The latter claim is then cashed out in modal terms by holding that there are relevantly close possible worlds where we have false moral beliefs. This received view, therefore, holds that debunking arguments undermine moral beliefs by taking the lack of an explanatory connection to indicate a lack of modal co-variation. In the end, it is the latter, modal profile of a belief that does the heavy lifting in the epistemological principles that underwrite debunking arguments on this view.

In this chapter, we will look at attempts to reorient debunking arguments toward employing non-modal, explanatory constraints of the sort Harman relied on. The type of explanationist arguments considered seeks to show that the lack of an explanatory connection between a set of facts and the beliefs they are about can undermine those beliefs *directly*, without merely being an indicator of a lack of appropriate modal-covariation.

If successful, such constraints could avoid the need to rely on modal conditions at all. This would allow a debunker to circumvent the troubles with counterfactuals and modality encountered in the previous chapter, while also managing to block the accidental correlation strategy which has tied its fate to the centrality of modal conditions for knowledge and justification.

Recall the explanatory constraint introduced in §3.3.4.

> EXPLANATORY CONSTRAINT
> If S' belief that p is not explained by the fact that p, or we lack an account of this explanatory relation, then S' belief that p is undermined.

Explanatory constraints of this sort have recently been reintroduced, refined, and defended in debunking contexts.[1] Unlike a causal constraint, the explanatory relation need not be causal. Rather, it must be *becausal*, in the sense that either the facts or the belief obtain *because of* the other.[2] This distinguishes it from modal constraints, such as sensitivity and safety, which only concern modal co-variation and where there need be no such explanatory relationship.

[1] Lutz (2018; 2020); Korman & Locke (2020; 2021). Faraci (2019), while not explicitly defending such a constraint as a defeater, lays the groundwork for such an application.

[2] Cf. Lutz (2020, sec. 13.4).

Despite their apparent attractions, I will argue that appealing to explanatory constraints in order to formulate global moral debunking arguments is, as it stands, unsuccessful. I will argue that explanationist proposals face two related problems. First, debunking arguments appealing to explanatory constraints seem to generalize in a way that threatens the justificatory status of our beliefs not just about morality, but also about logic, mathematics, modality, normativity, and even some empirical domains. Thus, if such arguments are sound, they would concede victory to a wide-ranging skepticism—in all likelihood much wider than what the debunker set out to defend.

Second, explanationist constraints risk being self-defeating since it is not obvious that our beliefs about explanationist constraints possess the relevant explanatory connections those constraints require. If our belief in the constraints do not satisfy them, any belief in such conditions are undermined.

I will return to these issues after having set out a recent explanationist proposal in some detail in the next section.

7.3 Korman and Locke's Explanatory Constraint

Daniel Z. Korman and Dustin Locke have recently argued that there is an explanatory constraint on rational belief.[3] They apply this constraint to the moral domain, where they, very roughly, propose the following principle: holding that one's beliefs about morality are not explained by, nor themselves explain, the moral facts is sufficient for those beliefs to be undermined.

Before moving on to discussing the details of Korman and Lock's proposal, it is helpful to adopt three of their technical terms. These describe two types of connections that can obtain between facts and beliefs, as well as an attitude one might take to the absence of such an explanatory relation.

First, moral facts can explain moral beliefs, or vice versa; in either case, the moral beliefs are then said to be explanatorily connected, or *e-connected*.

> E-CONNECTED
>
> One's moral beliefs are e-connected iff$_{\text{def}}$ moral facts either explain or are explained by one's moral beliefs.[4]

[3] Korman and Locke (2020). Korman and Locke do not explain their use of "rational belief." For many epistemic internalists, 'rational belief' is fully, or at least roughly, interchangeable with 'justified belief'. For the purposes of this chapter, I will treat them as interchangeable.

[4] Korman and Locke (2020, 310).

More generally, and roughly, a belief is e-connected “iff it explains or is explained by the sorts of facts it purports to be about.”[5] The rough idea being that we often believe that p *because* p, or that p is the case *because* we believe that p. I now believe that it is raining outside precisely *because* it is raining outside. Korman and Locke make no explicit restriction or specification of how ‘explanation’ is to be understood in this context, but for their argument to be plausible they must mean that the relevant facts at least partly explain the belief. The constraint is not meant to be restricted to causal explanation.[6]

Sometimes, p is the case, at least in part, because we believe that p. For example, the fact that paper bank notes hold monetary value (i.e., can be exchanged for goods and services) can be explained, at least in part, by it being a widely shared belief that bank notes hold monetary value. If no one believed that bank notes had monetary value, it would not be possible to exchange them for goods and services.

Other times, we might believe that p, but come to think that our belief is explained neither by p, nor that the belief that p helps explain why p is the case. In such circumstances, Korman and Locke claim that we are rationally committed to denying that our belief that p is e-connected (or, at least, to withholding judgment about whether our belief that p is e-connected). When one denies (or withholds belief about) a belief being e-connected in this way, they call this making an *explanatory concession* about the relevant belief.

When our beliefs are e-connected, this tends to license the inference that there is an important modal relationship between the relevant facts and beliefs. For instance, if the reason you think there is a flowerpot on the floor is because of causal impingement on your sensory apparatus—a causal, explanatory relation—then you would not have that belief if the flowerpot was not there (in close possible worlds, at least). An explanatory connection between a fact and a belief therefore tends to indicate an appropriate modal co-variation.

This leads us to the second type of connection that can hold between a belief and a fact, which is modal in nature. Sometimes, our beliefs and the facts they are about are modally related in an epistemically privileged way, such as being sensitive or safe. We can then say that the beliefs are modally connected, or *m-connected*.

[5] Korman and Locke (2020, 316 fn.15).

[6] Korman and Locke (2020, 324).

> M-CONNECTED
>
> One's moral beliefs are m-connected iff$_{\text{def}}$ one's moral beliefs bear some epistemically significant modal relation to moral facts.[7]

Here, "some epistemically significant modal relation" is a placeholder for whatever modal relation we think is required in order to make it rationally defensible to maintain our beliefs. One can therefore plug in one's favorite modal relation, whether it be "safety, sensitivity, reliability, and non-accidental accuracy," or some other relation.[8]

With these two kinds of connections on the table, Korman and Locke go on to claim that the typical view of debunking arguments against moral beliefs has roughly the following structure:

(1) Realists are rationally committed to believing that their moral beliefs are not e-connected.

(2) If one is rationally committed to believing that one's moral beliefs are not e-connected, then one is rationally committed to believing that one's moral beliefs are not m-connected.

(3) If one is rationally committed to believing that one's moral beliefs are not m-connected, then one is rationally committed to withholding from moral beliefs.

(4) So, realists are rationally committed to withholding from moral beliefs.[9]

Realists, who wish to block the debunking argument, can deny any of the three premises (1)-(3). Different types of realists will object to different ones. Certain non-naturalist realists will deny (1), by claiming that our moral beliefs are e-connected in virtue of the role moral facts play in explaining our moral attitudes.[10] Naturalist realists tend to reject (1) by claiming that the moral facts, in virtue of their relation to natural facts, can explain our moral beliefs.[11] Certain moral realists—such as non-naturalists—cannot so easily reject (1), since,

[7] Korman & Locke (2020, 316).

[8] Korman & Locke (2020, 310).

[9] Korman & Locke (2020, 311).

[10] When the possibility of such (non-causal) explanatory relations are suggested, they tend to be promissory notes rather than suggestions of mechanisms (e.g. W. J. FitzPatrick 2015, 894).

[11] Copp (2008); Lott (2018).

as we have seen at length in previous chapters, they have no available account of the explanatory relation between moral beliefs and moral facts, and so must make an explanatory concession concerning moral beliefs. This is what we have seen proponents of the accidental correlation strategy explicitly do.

Korman and Locke dub anyone granting the first premise but objecting to one of the others—usually premise (2)—'minimalists’. The minimalist response is so-named because it aims to neutralize the dialectical force of debunking arguments without introducing much, if any, controversial or additional theoretical machinery. As proponents of the minimalist response see it, they can grant the debunker that our beliefs about morality (or, indeed, any other subject matter) neither explain nor are explained by the moral facts, but without thereby being rationally committed to giving up on the idea that moral beliefs are m-connected.[12] According to Korman and Locke’s classification, both the third-factor strategy and the accidental correlation strategy would count as minimalist responses in virtue of their acceptance of (1).[13]

The minimalist umbrella comprises quite distinct strategies, and the details of how the minimalist response is spelled out differ depending on the beliefs and modal relations one considers. But the general form of the response should be familiar from the foregoing chapters. Consider how a minimalist who subscribes to the accidental correlation strategy might argue that our moral beliefs are likely to be safe.[14] There are three steps. First, the propositional content of our moral beliefs (or some subset thereof) will be argued to be non-accidentally true, in the sense that it is true in all close possible worlds.[15]

Second, it is argued that we have no reason to believe that moral beliefs are accidentally held, i.e., that we might easily have had different, and false, beliefs in close possible worlds. And lastly, it points out how it follows from the first and second steps above that the safety of our beliefs is not challenged by debunking arguments, in the sense that we have not been given any reason to think that we are wrong about what the moral facts are in any close world, and

[12] Discussion or defense of some form of minimalist response are found in Nozick (1981), Huemer (2005), Schafer (2010), Enoch (2010; 2011), White (2010), Wielenberg (2010; 2014), Brosnan (2011), Parfit (2011), Skarsaune (2011), Berker (2014), Clarke-Doane (2015; 2016; 2020), Talbott (2015), Vavova (2015), Baras (2017a), and Moon (2017).

[13] Cf. Korman and Locke (2020, 312).

[14] Korman and Lock take S’s belief that p to be safe if and only if S could not easily have been wrong about whether p. We could substitute the refined version of safety introduced in the previous chapter. Recall, SAFETY*: S’s belief that p, formed via method M, is safe if and only if, in every close possible world where S believes that p (or some sufficiently similar proposition p*), via method M, p (or p*) is true.

[15] For instance, by arguing that the fundamental moral truths are metaphysically necessary, in which case they would, of course, also be true even in distant possible worlds.

that this is the case regardless of whether or not our moral beliefs are e-connected.

Thus, proponents of the minimalist response claim that we can defend ourselves against the debunker by granting that our beliefs about morality are not e-connected, while still maintaining that they are m-connected.[16]

The type of minimalist response outlined above relies on the assumption that explanatory concessions only undermine the justification for the set of beliefs in question *indirectly,* by showing that the beliefs are not m-connected. But, according to Korman and Locke, this account of how explanatory concessions undermine is misguided. As they see it, it is not just that explanatory concessions *can* undermine directly, without having to first show that the beliefs whose justificatory status is in question fail to satisfy some epistemically significant modal relation, but that they *always* do so. Indeed, they tell us that

> explanatory concessions [...] undermine beliefs directly, and it is not in virtue of revealing the beliefs to be unsafe or unreliable or in some other way deficient that the concessions undermine those beliefs. […] [I]f explanatory concessions defeat directly, then the minimalist gambit can't get off the ground.[17]

To motivate their claim, Korman and Locke provide several thought experiments with the purpose of eliciting the intuition that explanatory concessions defeat the justificatory status of some target set of beliefs *even though* those beliefs satisfy whatever modal relations one thinks are epistemically significant. This means that in addition to cases where one can have unsafe or insensitive knowledge, mentioned in Chapter 6, Korman and Locke argue that there can be beliefs that, if true, *are* safe and/or sensitive, but which are nonetheless epistemically undermined. I will set out one of the scenarios they present, which targets safety.

> JACK
>
> Jack sees a streak in a cloud chamber and believes that the streak was caused by a proton. But Jack has not received the training of an ordinary physics student. Rather, he believes it because some Martians—after convincing him of their superior intellect—told him that protons cause those kinds of streaks. Moreover, they decide to tell him this, not because they themselves had done any physics, but simply because they liked the sound of the English word "proton." You may even suppose, if you like, that there is some deep law of Martian psychology that makes them like the sound of the word "proton," and so it could

[16] Clarke-Doane (2016) exemplifies this strategy, although he doesn't argue that our moral beliefs are, in fact, m-connected.

[17] Korman & Locke (2020, 317).

> not easily have happened that the Martians told Jack that such streaks were caused by something else. Finally, let us suppose that after forming the belief that protons cause those streaks, Jack learns all these details about the origins of his beliefs, and concedes that his belief that the streaks are caused by protons is not explained by the facts about what causes them.[18]

Two things need to be noted about this case. First, Jack's belief that the streak he sees is caused by a proton is not e-connected. It is clear from the description of the case that his belief neither explains nor is explained by the facts it is about. Second, it is compatible with the description of the case that his belief is m-connected. Indeed, not only is it clear that it could not easily have been the case that the streak was not caused by a proton, since it actually was, and the fact that it was is underwritten by natural laws. It also could not easily have been the case that Jack held some incompatible belief about the cause of the streak, since it is a fact about Martian psychology that they have a natural liking for the word 'proton'.

After learning what had happened, Jack makes an explanatory concession, in the sense that he now knows his belief to not be e-connected, even though it is (or at least might very well be) m-connected (by being safe). Nevertheless, despite being m-connected, Jack's belief does appear to be irrational given what he has learned. And, according to Korman and Locke, the best explanation for why the belief is irrational is that Jack's explanatory concession (all by itself) undermines it.

Korman and Locke think that similar cases can be constructed for other modal relations.[19] In this way, they argue against the minimalist response by rejecting the assumption that explanatory concessions can only defeat indirectly, by first showing that the target beliefs are not m-connected. Instead, they claim, intuitive judgments about scenarios such as JACK suggest that we take explanatory concessions to have the power to undermine *directly*.

There is of course the alternative possibility that the subjects in scenarios like JACK fail to satisfy some *other* modal relation, besides safety or sensitivity, and that this could be the explanation for our intuitions about those scenarios (namely, that the subjects are unjustified in holding their beliefs). However, Korman and Locke dismiss this option as a less "natural" explanation for the aforementioned intuitions.[20] They go on to argue that a more natural, and in fact the best, explanation of our judgments about cases like JACK, is that something like the following explanatory constraint is true.

[18] Korman & Locke (2020, 320–21). The scenario is originally from Locke (2014).

[19] For a similar scenario against sensitivity, see Korman and Locke (2020, 317–19).

[20] Korman and Locke (2020, 323).

> Explanatory Constraint* (EC*)
> If p is about domain D, and S believes that her belief that p is neither explained by nor explains some D-facts, then S is thereby rationally committed to withholding belief that p.[21]

A debunking argument constructed around EC* will differ from many of the arguments we have discussed so far. For one thing, it questions the assumption that explanatory concessions can only undermine beliefs indirectly, through targeting m-connections. This assumption has arguably been shared by both proponents of the minimalist response as well as debunkers who endorse the received view about how debunking arguments work ((1)-(4) above). An upshot of Korman and Locke's argument is that the received view too must be false (or, at best, misleading). In light of this, they offer the following, revised conception of how the structure of debunking arguments should be understood:

(1*) Realists are rationally committed to believing that their moral beliefs are not e-connected.

(2*) If p is about domain D, and S believes that her belief that p is neither explained by nor explains some D-facts, then S is thereby rationally committed to withholding belief that p.

(3*) So, realists are rationally committed to withholding from moral beliefs.[22]

Korman and Locke's revised formulation of their argument is not deductively valid as it stands. For one thing, the argument has an unstated premise that realists *in fact* possess the requisite second-order attitudes about explanatory concessions. Though not made explicit by Korman and Locke, that means that EC* only applies to agents who hold second-order beliefs about the e-connectedness of a given belief. EC*, as it stands, will therefore not undermine the beliefs of anyone who lacks such second-order beliefs. In the case of moral beliefs, most non-metaethicists presumably lack such second-order attitudes. As we will see, this results in a debunking argument employing it being restricted in its target.

[21] Korman & Locke (2020, 325).
[22] Korman & Locke (2020, 327).

On this revised understanding of how debunking arguments can harness the lack of explanatory connections between moral beliefs and moral facts, there is no room for the type of minimalist reply outlined earlier. Since the minimalist's acknowledgment of a lack of e-connection undermines the relevant beliefs directly, there is no further question of whether such a theorist could reason their way to the safety and sensitivity of those beliefs. Whether or not an undermined belief has a certain modal profile or other is a moot point.

This severely limits the type of replies that are available to moral realists, and in particular to those who would normally accept premise (1) and deny premise (2) of the received construal of debunking arguments. If EC* is true, therefore, it would imply a radical refiguring of the available responses to evolutionary debunking arguments.

We have now seen how Korman and Locke argue that it is not a failure to satisfy some modal condition that is the explanation for how debunking arguments undermine beliefs. Rather, the claim is, the lack of explanatory connections undermines directly. How plausible is this claim, taken on its own? Korman and Locke support the claim by setting out scenarios like JACK, where m-connection is secured, but where they claim the agent's beliefs are still undermined. Responding to Korman and Locke, Clarke-Doane flat-footedly rejects the importance of a lack of e-connection insofar as m-connection is secured.

> Some might still feel that *connection* per se matters, even if it is not predictive of reliability in any useful sense […]. And I suppose that P's being implied by some explanation of our coming to believe that P, even for causally inert P, is *some* kind of connection between P and our (token) belief that P. But, again, while we are free to use "coincidence" to satisfy the condition that P is true by coincidence if all explanations of our coming to believe that P fail to imply that P, there seems to be no epistemic reason to *care* about coincidences, so conceived.[23]

Clarke-Doane's claim is that *if* we can show our beliefs to be non-accidentally true, and that they could not easily have been false, it is not clear why we should worry about a lack of e-connection. Clarke-Doane claims that we should be perfectly happy if we can check all boxes *other* than e-connection.

> [F]rom anything resembling the standpoint of trying to have true beliefs, it is surely preferable to have a safe, sensitive, and probable belief which is not "connected" to the truth than to have an unsafe, insensitive, and improbable belief that is.[24]

[23] Clarke-Doane (2020, 119), emphasis in the original.

[24] Clarke-Doane (2020, 119).

The claim made by Korman and Locke, of course, is not that beliefs that are e-connected *without* being m-connected necessarily have a positive epistemic standing. Rather, it is that beliefs that are m-connected without being e-connected are necessarily undermined. Unlike Clarke-Doane, I am sympathetic to the notion that a lack of explanatory connections *does* matter, epistemically, even if m-connections could hypothetically be established.[25]

That being said, I believe that the debate over the plausibility of the direct undermining force of explanatory concessions has all the hallmarks of a disagreement in intuitions that is unlikely to be resolved without being able to appeal to independent evidence. In the following two sections, I will therefore argue that an explanatory constraint like EC* faces two major obstacles, independently of the plausibility of whether explanatory concessions undermine directly. First, I argue that such principles are faced with the threat of overgeneralization. Then, in §7.5, I show that EC* is *prima facie* self-defeating.

7.4 Do Explanatory Constraints Overgeneralize?

In this section, I argue that explanatory constraints such as EC*, if true, threaten to undermine beliefs that it is generally accepted that it would be perfectly rational for us to believe. Hence, such constraints are not extensionally adequate. I will, for the most part, present my arguments as targeting EC* in particular, but much of my discussion should apply to other explanatory constraints as well. For instance, the arguments in this section and the next will, at least for the most part, also apply to the EXPLANATORY REQUIREMENT discussed in §2.2, as well as EXPLANATORY DISPENSABILITY and EXPLANATORY CONSTRAINT, discussed in §3.3.2 and §3.3.4, respectively.

When it comes to extensional adequacy, there are two separate problems facing EC*. On the one hand, EC* threatens to undermine beliefs in mundane empirical propositions, in particular beliefs about the future and inductive knowledge. On the other, it threatens to undermine a wide swath of philosophical, mathematical, and other a priori beliefs.

That EC* threatens to undermine certain mundane beliefs about empirical propositions that we take ourselves to have good reason to hold is taken up by

[25] One reason for this sympathy stems from the possibility of constructing what Faraci (2019) calls same-modality contrast cases. These are pairs of cases, where both satisfy the relevant m-connection, but where one case nonetheless seems to be afflicted by a malignant epistemic coincidence while the other does not. Faraci identifies the malignant coincidence with the absence of an explanatory connection.

Korman and Locke in passing. They consider the belief that the sun will set in the west.[26] Is that belief e-connected? That depends on how one classifies which domain the belief is about—is it about *the sun, the west, the future*, or all of them?

The belief will be e-connected as long as it is taken to be about the sun and/or the west. Certain facts about the sun and the west plausibly enter into the explanation of our belief that the sun sets in the west. It would therefore not run afoul of EC*. If the belief is taken to be about the future, then the belief is not e-connected, since it neither explains nor is explained by the fact that the sun will set tomorrow (or, indeed, by any other fact about the future). Moreover, since a similar line of reasoning can be used against all other beliefs about the future, EC* appears to lead to skepticism about the future. Beliefs about the future cause trouble for explanationist theories in general, so it should not be surprising that the issues turn up here as well.[27]

The viability of EC* therefore hangs on how we resolve the issue of what domain(s) a belief is about. Korman and Locke seem to suggest that the debunker should avoid such skepticism about the future by gerrymandering which domains beliefs are understood to be about.

> [T]he proponent of EC* must supply some account of which domains are relevant to assessing whether a belief satisfies EC*, one which excludes *the future*.[28]

This might seem questionably *ad hoc*, although I will not pursue that issue here. Absent a solution to this type of "generality problem" of domain individuation, EC* looks poised to undermine beliefs that it would seem perfectly rational for us to maintain belief in. Arguably, failing to allow for justified belief about the future shows that EC* is extensionally inadequate.[29]

Perhaps some variant or reformulation of the explanatory constraint can be formulated which avoids the above problem.[30] In a recent paper, Korman and Locke jettison EC* because of the trouble it runs into with respect to individuating domains. There, they opt for a different principle.

[26] Korman and Locke (2020, 325).

[27] See McCain (2015) for an attempt to defend a general explanationist theory of epistemic justification from similar worries.

[28] Korman & Locke (2020, 325), italics in the original.

[29] Of course, one can always choose to bite the bullet and accept skepticism about the future. To most, I suspect, this will instead constitute a *reductio* of the view.

[30] They introduce one such variant, which would allow third-factor explanations to satisfy it, but think that it falls prey to counterexamples (Korman and Locke 2020, 325–26).

> E<span>REASONS</span>
> If S is not entitled to believe that the facts she treats as reasons to believe that p support* her belief that p, then S's belief that p is defeated.[31]

There is quite a bit to unpack in this principle. I will only do so briefly. For Korman and Locke, being *entitled* to hold a belief requires being rationally permitted to believe it. For a fact to *support** a belief, the fact must do two things. It must logically support the belief, either deductively or inductively. It must also do so, *relative to all other things that the agent ought to take to be the case*.

Given this, the idea is that when we believe that p, we need to be epistemically entitled to thinking that whatever reason we have to believe that p logically supports the truth of the content of that belief. Furthermore, we must take this to be the case given the totality of what we ought to take to be the case.

Applied to facts about the future, Korman and Locke state that EREASONS succeed where EC* failed. We are entitled to believe that the facts (about the past and present), which we treat as reasons to hold beliefs about the future, supports* such beliefs. Facts about past sunrises, for instance, inductively support our belief in future sunrises given the totality of what we ought to believe.

Korman and Locke do not discuss this new principle in connection with moral debunking, but we can apply it. Consider this from the viewpoint of a moral non-naturalist adherent of the accidental correlation strategy. Such a non-naturalist, we can grant for the sake of argument, is defeasibly justified in having the belief that it was wrong of Paul to unnecessarily cause John pain.[32]

Will EREASONS cause moral beliefs like this to be undermined? We saw that EC* would straightforwardly undermine this belief, at least insofar as the non-naturalist would make an explanatory concession with respect to it. To answer this question, we must first ask what facts an agent could have for holding the belief. Assume that Paul did in fact cause John unnecessary pain. A non-naturalist could plausibly hold, again with defeasible justification, that causing unnecessary pain is bad.

EREASONS would undermine the belief that it was wrong of Paul to unnecessarily cause John pain if the agent is not entitled to believe that the reasons she has for believing it support* it. Given that Paul caused John unnecessary pain, it would seem to follow that this act was wrong. The non-naturalist therefore seems entitled to believe that the fact that causing pain unnecessarily is wrong

[31] Korman and Locke (2021, 15), emphasis removed.
[32] Korman and Locke themselves are happy to grant it.

supports* the belief that it was wrong of Paul to unnecessarily cause John pain. This, furthermore, is so given the totality of what the non-naturalist should take to be the case. As such, $\text{E}_{\text{REASONS}}$ would not seem to pose any threat to the non-naturalist's moral beliefs.

While $\text{E}_{\text{REASONS}}$ seems capable of avoiding the problems connected to beliefs about the future, it is not clear how it is supposed to undermine the non-naturalist's defeasibly justified moral beliefs. The reason the undermining power of the principle has evaporated is that $\text{E}_{\text{REASONS}}$, on its face, is not an explanatory constraint. It poses no requirement that beliefs be e-connected. Korman and Locke do, however, attempt to turn $\text{E}_{\text{REASONS}}$ into an explanatory constraint by defending the following ancillary principle that they call *Treating Requires an Explanatory Connection* (TREC).

> TREC
> If the fact that p is not part of what explains S's belief that q, then S does not treat the fact that p as a reason to believe that q.[33]

TREC holds that one can only treat a fact as a reason for belief if the fact is part of the explanation of the agent's belief. By adopting TREC, $\text{E}_{\text{REASONS}}$ turns into an explanatory constraint. This is because TREC constitutes a necessary condition for treating a fact as a reason. If we again consider a non-naturalist proponent of the accidental correlation strategy, they would be barred from ever treating any fundamental moral fact as a reason, as those facts are not part of the explanation of any of our beliefs. As such, with TREC in place, $\text{E}_{\text{REASONS}}$ could again succeed in undermining moral beliefs.

At this juncture, the heavy lifting is being done by TREC. How plausible is this principle? Discussing its plausibility at length would take us too far afield, but let me register a few concerns. Consider a case where an agent acts on the belief that p, but where p is false. Either we are allowed to take agents to treat such false propositions as reasons, or we are not.

If we are not, TREC seems to diverge strongly from how we would ordinarily describe what agents do. In ordinary parlance, one often gives as a reason, and, one would think, *treats as a reason,* things that are not facts at all. It is unclear how TREC would handle such cases. If we cannot treat false propositions as reasons, this leads to questions about the linguistic or metaphysical legitimacy of requiring explanatory connections in order to treat something as a reason.

[33] Korman and Locke (2021, 15).

If, on the other hand, we *are* allowed to treat as a reason things that are not facts, it would seem strange to disallow treating explanatorily unconnected facts as reasons. Then we could treat a false proposition P as a reason, but not the true but explanatorily disconnected fact that Q.

The issue of whether an explanatory concession—an acknowledged lack of, or withholding of belief about, e-connection—undermines directly is plausibly still the central issue. But it has now been relegated to an ancillary principle. And that principle does not outright claim that lack of an e-connection undermines directly, so much as it seemingly rules out the possibility of treating explanatorily disconnected facts as reasons. For reasons of simplicity, I will continue the discussion in terms of Korman and Locke's original principle, EC*. Even so, I will sometimes relate the discussion to TREC and EREASONS as well.

Moving beyond beliefs about the future, there is a second and related challenge arising from concerns about generalization for explanatory constraints such as EC* and TREC. While Korman and Locke only apply EC* to the moral domain, and consider its potential generalization to facts about the future, there are other domains where we are no less likely to make explanatory concessions.

Consider mathematical beliefs. If we are mathematical Platonists, entities such as numbers are abstract, non-spatial, and non-causal. If anything, such claims are even more plausible about mathematical entities such as numbers than about moral facts.[34] Given that this is so, the question becomes whether, say, our belief that 2 is prime is either a part of the explanation of the fact that 2 is prime, or if the belief is explained by that fact.[35]

It is implausible to think that your belief in the mathematical fact explains why it obtains (or, indeed, why any other mathematical fact is true). On the other hand, it is hard to see how the mathematical facts could explain our belief that 2 is prime. If we frame this in terms of explaining the correlation between our mathematical beliefs and the mathematical facts—analogous to the MINIMAL CORRELATION from §5.3.1—we can see this more clearly. It is unlikely that our beliefs stand in some direct relationship to the mathematical facts, either by our beliefs explaining the facts or the facts standing in some direct causal or non-causal relationship to our beliefs.

Secondly, EC* seeks to rule out indirect explanations of the form employed by third-factor explanations. This leaves the accidental correlation strategy

[34] Cf. Clarke-Doane (2020, 73).

[35] Similarly, if considering TREC, the question is whether mathematical facts are part of the explanation of our mathematical beliefs. If not, we could not treat such facts as reasons.

and full-blooded Platonism. As EC* rules out the accidental correlation strategy, the mathematical Platonist might find themselves in much the same situation as the moral non-naturalist.[36]

This line of argument shows that a defender of explanationist principles cannot content themselves with debunking moral non-naturalist realism. Their arguments will likely have a further reach, which will mean that a debunker employing it will have to be prepared to go beyond the moral domain.

One might think that mathematics will nonetheless be able to satisfy principles such as EC* and TREC in a way that morality may not—by appealing to indispensability. Consider the claim by Mark Steiner that any physical explanation whatsoever will need to appeal to the axioms of number theory.[37] If so, any (physical) explanation of our mathematical beliefs (or of any other belief), therefore includes the axioms of number theory. As such, one might think that, at the very least, a belief in the axioms of number theory could be e-connected (in this trivial sense).

While this might seem to render mathematical beliefs trivially explanatorily connected, it is not so straightforward. As Korman and Locke themselves acknowledge, while some mathematical facts—i.e., the axioms of number theory—could be thought to secure such a relation, it is not a given that this applies to all other mathematical facts. Recall that TREC requires that for *any* fact that p, to treat that fact as a reason for believing that q, p needs to be part of the explanation of your belief that q. Whether this can be accomplished for a sufficiently wide array of mathematical facts is not settled merely by the axioms of number theory being trivially entailed by the explanation of any physical phenomena whatsoever. The danger of EC* (or TREC) rendering belief in many mathematical facts irrational is therefore still a live worry.

The above concern also applies to other domains, including metaphysics, logic, and modality. It can be hard to see how metaphysical principles of composition or ways the world could have been explain our beliefs about such facts.[38] In short, the premises of Korman and Locke's revised argument—(1*)–(3*)—would seem to be equally forceful when targeting traditionally a priori domains. While it is open to the debunker to accept this consequence, and thereby opt to launch a very wide debunking argument targeting all, or much of, a priori knowledge, this would constitute a much broader debunking argument than most debunkers would be willing to defend.

[36] An interesting question, which we will not investigate here, is what an explanatory constraint such as EC* would say about mathematical belief as conceived by the full-blooded Platonist.
[37] Steiner (1973, 61–62).
[38] For relevant discussion of metaphysical principles of composition, see §6.3.3.

While cursory, the above discussion is surely enough to put pressure on the debunker to be faced with the following choice: Either a debunker employing EC* must *deny* that (or withhold judgment about whether) the above beliefs—e.g. beliefs about the mathematics, metaphysics, modality, logic, and other similar domains—are e-connected, or she must provide some reason to think that such beliefs *are* e-connected.

If the debunker chooses the first option, then EC* grounds an extremely broad debunking argument. If the debunker instead opts for the second option, they take on the burden of having to provide reasons to believe that our beliefs about mathematics, metaphysics, modality, logic, and other similar domains are e-connected.[39]

This leads to a major dialectical shift which Korman and Locke do not seem to recognize. In the absence of an account of why this or that domain is e-connected, skepticism takes hold. At least this holds true as long as the debunker has considered the issue and holds the requisite second-order concessive beliefs. And so, by embracing EC*, and being presented with the threat of generalization, a debunker takes on an enormous explanatory burden.

At this point, it is worth noting the awkward position the threat of generalization puts the debunker in. The burden of proving that one's beliefs about a large number of domains are e-connected has now fallen *on the debunker*. The debunker, starting out with the goal of holding moral realists accountable for not being able to show that their beliefs are e-connected, now finds herself in the same position with respect to a host of other types of non-moral beliefs.

Furthermore, if the debunker were to develop some account that shows how our beliefs about, say, mathematics, logic, or metaphysics are in fact e-connected, those explanatory models could reasonably be expected to apply to the moral domain as well. And so the debunker, in virtue of defending her debunking argument, could come to serve up the very method by which a moral realist theorist could rebut it. While this mere possibility does not in and of itself speak against the truth of EC*, it does pose a dialectical obstacle to using EC* as the operative principle in a debunking argument.

So far, the threat of generalization has been a serious, but not necessarily lethal, problem. In the next section, we will see why it very well could be.

[39] A third option might be to claim that EC* is not generally applicable, and hold that it does not apply to, say, mathematics or logic. But this would seem objectionably *ad hoc*.

7.5 Are Explanatory Constraints Self-Defeating?

We have seen that explanatory constraints, such as EC*, risk generalizing in such a way that it threatens the justification we ordinarily take ourselves to have for empirical, mathematical, and philosophical beliefs. The second, more serious problem facing a principle like EC* springs from the first one. It is that EC* *itself* appears to be located within an a priori subject matter that could very well be targeted by EC*—namely, epistemology. If EC* generalizes in such a way that it undermines belief in a priori propositions, and thereby epistemological beliefs like a belief in EC*, then that threatens to undermine any justification one might have for believing in the principle itself. EC* is therefore *prima facie* self-undermining.

When setting out a debunking argument employing EC* (or some variant thereof) as its operative epistemological principle, a debunker should be able to rationally endorse that principle. A debunking argument will be self-defeating whenever the same epistemic defect it attributes to its target beliefs attaches to a debunker's own belief in the operative epistemological principle. The debunker is then barred from rationally endorsing the principle *by their own lights*. This is the type of epistemic self-defeat that threatens EC*.

To see why EC* is threatened by self-defeat, let us say that Korman and Locke's argument, presented above, is sound and that EC* is true. They must then either hold their belief in EC* to be e-connected, or not. If not, then according to EC* itself, their epistemic belief in the principle is undermined. Hence their argument is self-defeating. If Korman and Locke hold that EC* *is* e-connected, the question becomes *how*? Korman and Locke provide no reason to think that epistemological principles like EC* are e-connected. Despite how critical it is to show that principles such as EC* can avoid self-defeat, Korman and Locke provide no discussion of that issue.[40]

Elsewhere, Korman in fact claims that debunkers are likely to *deny* that epistemic facts of this kind are e-connected. When discussing an epistemological principle much like EC* in an overview of the debunking literature, Korman writes that it "looks to be precisely the sort of abstract, normative fact that according to debunkers doesn't or can't explain our beliefs."[41]

If EC* is not e-connected, then debunking arguments making use of it are self-defeating. Insofar as many, and perhaps especially debunkers, doubt that

[40] In their more recent paper, they explicitly refrain from discussing issues of self-defeat (Korman and Locke 2021, 3 fn. 3).

[41] Korman (2019a, 8).

principles such as EC* are e-connected, the type of debunking argument presented by Korman and Locke is *prima facie* self-defeating. While worrying, this need not be the end of the road for debunking arguments based on explanatory constraints like EC*. In the rest of this section, we will explore the prospects for avoiding self-defeat for a debunker employing EC*.

For a debunker to be able to employ EC* without suffering self-defeat, they must show that EC* is e-connected. To see how it could be, one can start by considering the strategies that Korman and Locke suggest that moral realists can employ in order to show that moral beliefs are e-connected.

> Reductive views on which the moral facts just are the very natural facts that ultimately explain our moral beliefs are still in the running. So are theistic views on which the moral facts influence our moral beliefs by way of making themselves known to an intelligent designer who ensures that evolutionary processes yield reliable moral faculties. So are rationalist views on which moral facts influence our moral beliefs via some sort of quasi-perceptual apprehension.[42]

While Korman and Locke only concern themselves with realist views, it is incumbent on non-skeptical anti-realist views to show that moral beliefs are e-connected as well. To the above list we can then therefore add further possibilities, such as moral constructivism, which takes the explanatory relation to run in the opposite direction of the views outlined by Korman and Locke. On such views, it is our attitudes, practices, or agential natures that construct or constitute the moral facts.[43] If we look beyond cognitivist views, expressivist accounts of moral facts might also be able to secure an e-connection as well. *Mutatis mutandis*, these same strategies are available when attempting to show that epistemic beliefs are e-connected.

Here is the first lesson to be drawn from the discussion so far. A debunker cannot employ a constraint like EC* without committing to some such metaepistemological account that would allow it to avoid self-defeat. This is no small cost, as can be gleaned from Korman and Locke's comments on the list of metaethical realist options they outline.

> These all have problems of their own, to be sure. But they are the sorts of responses that are not ruled out by what we have shown.[44]

[42] Korman and Locke (2020, 326–27).

[43] For epistemic constructivist proposals, see Street (2009) and Warenski (2021, sec. 4).

[44] Korman and Locke (2020, 327).

Though they do not seem to recognize it, the same applies to their meta epistemological counterparts, which they would need to pick from in order to avoid self-defeat. By being forced to pick between the meta epistemological equivalents of the above metaethical views, a debunker is already forced to take on costly theoretical commitments. Those costs include explaining how epistemic beliefs—at least the belief in EC*—are e-connected.

To discharge such explanatory burdens, could a debunker merely pick and choose between available methods for establishing how epistemic beliefs are e-connected? As we will see, things are not so simple. There are in fact a number of further constraints on the metaepistemological strategies available to the debunker. At this point, it becomes imperative to be clear about the structural differences between different types of debunking arguments. This is particularly true of the distinction between what I in section §1.3 called non-conditional and conditional debunking arguments.

Non-conditional debunking arguments attempt to straightforwardly undermine the beliefs in the domains they target. Conditional debunking arguments, on the other hand, claim only that the beliefs of the target domains are undermined *given a particular metaethical conception* of that domain. For instance, Street argued that, given a moral realist construal of moral facts, moral beliefs are uniformly undermined. But she then goes on to argue that we should be metanormative constructivists (or non-skeptical anti-realists of some other stripe). I will now proceed to argue that each of these types of arguments face a challenge from the threat of self-defeat, but that these challenges importantly differ.

7.5.1 Non-Conditional Debunking Arguments and the Threat of Self-Defeat

Consider first non-conditional debunking arguments that rely on an explanatory constraint such as EC*. How can they respond to the threat of self-defeat? To show that their epistemic beliefs are e-connected, a debunker could endorse a moderate form of epistemic rationalism. For instance, they could attempt to explain the e-connectedness of epistemic beliefs through the claim that epistemic facts are conceptual truths.[45] One might think that the fact that knowledge requires true belief is a plausible candidate for being a conceptual

[45] I assume that such conceptual competency can be cashed out in a way that satisfies EC*. If not, so much the worse for this strategy. This line of argument might also apply to other theories of *a priori* knowledge that attempts to cash it out in terms of conceptual competency.

truth. Perhaps one thinks that it is part of the concept of ‘knowing that’ that, necessarily, if anything satisfies the concept ‘knowing that’, it also satisfies the concept ‘truly believing that’. And so on for other epistemic truths.

I do not intend to evaluate the plausibility of such a view. Whatever its merits, assume that a debunker used it to explain the e-connectedness of our epistemic beliefs. Such an attempt at securing the e-connectedness of epistemic beliefs face an obvious challenge. Whenever a debunker suggests a strategy for showing how epistemic beliefs are e-connected, that strategy is quite likely to be equally applicable to the moral domain as well.

Consider Terence Cuneo and Russ Shafer-Landau’s argument that the moral domain consists, at least in part, of non-natural conceptual truths.[46]

> There are nonnatural moral truths. These truths include the moral fixed points, which are a species of conceptual truth, as they are propositions that are true in virtue of the essences of their constituent concepts.[47]

Here is an example of what it would mean to be a conceptual truth of the kind they have in mind. That is, a conceptual truth where a proposition is true in virtue of the essence of the concepts occurring in it.[48]

> [C]onsider the proposition <that recreational slaughter of a fellow person is wrong>. This is a conceptual truth in case it belongs to the essence of the concept ‘being wrong’ that, necessarily, if anything satisfies the concept ‘recreational slaughter’ (of a fellow person) it also satisfies ‘being wrong’ (in a world sufficiently similar to ours).[49]

To the extent that one embraces such an account of epistemic truths, one would seem to provide motivation for adapting the same view about the truths of morality.[50] In fact, this type of moral-epistemic parity would seem to hold true of every form of rationalism that could be used to secure the e-connectedness of epistemic beliefs. It holds true, for instance, of stronger forms of rationalism that hold that epistemic beliefs are e-connected because we can

[46] Cuneo and Shafer-Landau (2014).

[47] Cuneo and Shafer-Landau (2014, 411–12). They elucidate the notion of the “essence” of concepts with an example: “The concept ‘being wrong,’ for example, could not be the concept it is if it were not about wrongness; it belongs to the essence of the concept that it applies to exactly those things that are wrong (if any such things there be)” (Cuneo and Shafer-Landau 2014, 410).

[48] I here gloss over many details of their view, such as the particular account of concepts they employ.

[49] Cuneo and Shafer-Landau (2014, 410).

[50] Kyriacou (2018) develops a detailed argument for this claim.

come to know the epistemic truths through some quasi-perceptual rational ability that puts us into contact with the epistemic truths themselves.[51]

This means that opting for any form of epistemic rationalism renders a debunker likely to open the way for an equivalent form of moral rationalism. That is so, at least, unless the non-conditional debunker establishes some form of moral-epistemic disparity. In the absence of some such disparity, this form of metaepistemic strategy risks generalizing and shielding the beliefs targeted by the debunking argument. While the debunking argument would therefore no longer target epistemic beliefs, and hence not be self-defeating for that reason, it would no longer successfully target moral beliefs either.

What started as an attempt at debunking a particular set of beliefs would then instead have provided protection for those beliefs. While not straightforwardly a case of epistemic self-defeat as I have defined it, it certainly renders the debunking argument *qua* debunking argument toothless and, in a different sense, self-undermining.

I believe that considerations similar to those I have set out for various versions of epistemic rationalism also hold true of the other options for securing the e-connectedness of epistemic beliefs. For instance, defending some kind of theist explanation with respect to epistemic facts but not moral facts would be puzzling. Although I will not argue for it here, I believe constructivism, expressivism, and reductive naturalism about the epistemic domain will, by default, be likely to generalize to the moral domain as well.[52] This is primarily because of the similarities between the epistemic and the moral domain.

This presumption might not carry much weight on its own, but there are more general reasons for taking it seriously. To see why, observe that it is widely agreed that there are significant similarities between the epistemic and moral domains. A popular view holds that moral and epistemic facts are sufficiently similar that their fates are intertwined; if one domain is subject to

[51] Cf. BonJour (1998). Note that more lightweight forms of epistemic rationalism, such as phenomenal conservativism, are not, on their own, sufficient for establishing e-connectedness, as they do not require any explanatory connections between the beliefs and the facts themselves.

[52] An interesting exception here is Heathwood (2009; 2018) who is sympathetic towards *moral* non-naturalism and even thinks that any form of analytical reductionism is implausible in the moral domain. However, he thinks that *epistemic* analytical reductionism is far more plausible. This is because he believes that the moral variety, but not the epistemic, falls prey to the Open Question Argument. For criticism of Heathwood's general proposal, see Rowland (2013), Cuneo and Kyriacou (2018), and Kyriacou (2019, sec. 5).

wholesale rejection, so is the other.[53] Cuneo has famously defended the following version of a moral-epistemic parity claim.

> PARITY
>
> If moral facts do not exist, then epistemic facts do not exist.[54]

PARITY is a claim about metaphysics, about what the facts are. Debunking, as I have understood it, restricts itself to making epistemic claims about what we have justification for believing, or when our beliefs amount to knowledge. Let us therefore translate the parity claim into epistemic terms.

> E-PARITY
>
> If our moral beliefs are uniformly undermined, then our epistemic beliefs are uniformly undermined.[55]

I will not defend either version of the parity claim, but to the extent that one finds the metaphysical formulation plausible, one is likely to find the epistemic formulation plausible as well.[56] The widespread acceptance and intuitive plausibility of parity claims result in a debunker being put in something of a bind.

A non-conditional debunker employing EC* is faced with two options in light of the proclaimed similarities between the moral and the epistemic domain. The first option is to hold that their debunking argument succeeds and that moral beliefs are therefore uniformly undermined, and accept that parity holds. If E-PARITY is true, and a debunker claims to have globally undermined moral beliefs, then the debunker would need to accept that they cannot rationally maintain a belief in the epistemological principle they employ as a premise in their debunking argument. This results in straightforward self-defeat.

The second option is to hold that moral beliefs are uniformly undermined but that E-PARITY is false. If a debunker denies E-PARITY, and is therefore able to show that our epistemic beliefs *are* e-connected, the question will be

[53] Some who subscribe to a version of this claim are Kim (1988), Shafer-Landau (2006), Ridge (2007), Chrisman (2007), Street (2009), Rowland (2013), Greco (2015) Kyriacou (2016), Case (2018), Cuneo and Kyriacou (2018).

[54] Cuneo (2007, 89).

[55] To be plausible, E-PARITY is subject to a number of qualifications and clarifications, just like PARITY is. As my aim is not to defend E-PARITY, I will gloss over those here.

[56] Note that the claim could even be strengthened, as the converse of E-PARITY seems at least as plausible; namely, that if our epistemic beliefs are uniformly undermined, then our moral beliefs are uniformly undermined as well.

why, given the similarities to the moral domain, moral beliefs are not e-connected. That is, a debunker will have to show that there is a relevant disparity between the moral and epistemic domains. This involves showing that whatever metaepistemological strategy the debunker uses to vindicate their epistemic beliefs, would not similarly be capable of shielding moral beliefs. This option avoids self-defeat, but at the cost of denying parity and providing a metaepistemological and metanormative picture of the moral and epistemic domains.

We can now draw a second lesson, which applies to non-conditional debunkers who attempts to avoid self-defeat. Such a debunker needs to show that whatever strategy they employ for the purpose of showing epistemic beliefs to be e-connected is not equally plausible when applied to moral beliefs. This will by necessity deny a widely accepted claim about moral-epistemic parity, and thus require a debunker to provide a disunified metanormative picture of the moral and epistemic domains.

My ambition in this section has not been to show that non-conditional debunking arguments employing explanationist constraints cannot succeed. What I take to have shown is that a debunker cannot, like Korman and Locke, endorse a debunking argument based on an explanationist principle without doing very substantial work far outside of metaethics. This work will involve spelling out one's metaepistemological commitments as well as denying widely accepted claims about moral-epistemic parity. In short, the cost of admission for running a successful non-conditional explanationist debunking argument has been severely underappreciated.

7.5.2 Conditional Debunking Arguments and the Threat of Self-Defeat

Let us now move on to consider how the conditional debunker who employs a principle such as EC* fares when faced with the threat of self-defeat. Recall that some debunkers employ conditional debunking arguments that have a conclusion of the following type.

> Certain metaethical theories (e.g., moral realism)—but not all such theories—should be rejected.

A debunker defending a conditional debunking argument is not arguing that *no* moral belief is justified. Separately from their conditional debunking argument, they defend some account of how our moral beliefs *are* justified. Consider Street's conditional debunking argument, discussed in Chapter 4. Street targets the normative realist, who holds normative facts to be stance-independent. She argues that *if we were realists*, then our moral beliefs would be uniformly undermined. Street does not think we should be moral realists, however, as she thinks we should be moral constructivists. Such constructivists, in Street's view, can show that moral beliefs are e-connected.

If a conditional debunker such as Street employed a principle like EC* in their debunking argument, they would need to show that it is e-connected as well. In virtue of running a conditional argument, such a debunker has options available for explaining the e-connectedness of epistemic beliefs that are not available to either moral realists nor to non-conditional debunkers.

Street herself defends constructivism not only about moral (as well as other normative) facts but about epistemic facts as well. According to epistemic constructivism, epistemic facts are grounded in our nature as agents, our practical standpoint, or otherwise "constructed" from our attitudes.[57] It is therefore an anti-realist view, in the sense that it denies that any epistemic facts are stance-independent. Even so, an epistemic constructivist might be able to show that our epistemic beliefs are e-connected. This is so since our attitudes are part of the explanation for the epistemic facts being what they are.

The fact that a conditional debunker has this two-layered position consisting of both an argument against moral beliefs *given a particular metathetical construal,* together with a positive, unified account of the moral and epistemic domains, allows them additional resources to rebut the threat of self-defeat.

Recall that in order to avoid self-defeat, a debunker must do three things; (a) explain why epistemic beliefs are e-connected, (b) show that the beliefs targeted by their debunking argument are not e-connected, and (c), show that their strategy for establishing (a) does not generalize so as to also shield the beliefs targeted. We saw that a non-conditional debunker runs into trouble by either failing to secure (a), in which case they suffer self-defeat, or failing to secure (c)—and therefore (b)—in which case they fail to debunk their intended target beliefs.

A conditional debunker like Street can explain (a)—how it is that our epistemic belief about principles like EC* are e-connected—by appealing to her positive, epistemic constructivist proposal. She can explain (b) by showing that *given a realist interpretation* of moral facts, they fail to be e-connected.

[57] For some constructivist proposals for the epistemic realm, see Warenski (2021, sec. 4) and Street (2009).

However, given Street's own unified metanormative constructivist view, moral facts are *in fact* e-connected. Lastly, it might seem as if Street runs into trouble with (c), because the strategy she employs for showing that her epistemic beliefs are e-connected—epistemic constructivism—generalizes so that it applies to moral beliefs as well.

This is not a problem for those who, like Street, only seek to argue that *given a realist conception* of moral facts, our beliefs about them would not be e-connected. That moral beliefs are e-connected *given metanormative constructivism* is in fact Street's desired outcome. This has the added benefit that a conditional debunker like Street, who presents a unified metanormative account of the moral and epistemic domain, will not have to deny E-PARITY, but is instead likely to endorse it.

This allows us to draw a lesson for the conditional debunker. A conditional debunker that commits to a positive, unified account of the moral and epistemic domains has a way of avoiding self-defeat without having to deny moral-epistemic parity. That advantage comes at the cost of needing to supplement one's debunking argument with a full-fledged, unifying metanormative account that covers at least the moral and epistemic domains. In Street's case, the ability of her debunking argument to avoid self-defeat therefore rests in large part on the plausibility of her metanormative constructivism.

This outcome should be surprising. It runs counter to the typical narrative of debunking arguments being skeptical arguments. If what I have argued above is correct, by far the most promising type of global moral explanationist debunking argument is one which *by its very nature* seeks to grant us true, justified beliefs about both the moral and the epistemic domain.

7.6 Conclusion

This chapter has investigated the prospects for debunking arguments employing explanatory constraints as sufficient conditions for epistemic defeat. We have seen that arguments employing explanatory constraints such as EC* face several challenges, which, if left unanswered, make debunking arguments employing it either unattractive or self-defeating. First, such arguments seem to rule out justified belief about the future. Second, they risk ruling out justified belief about many domains concerned with a priori knowledge, such as mathematics, metaphysics, modality, and logic. Third, one such arguably a priori domain is epistemology, which, if EC* undermines it, would make an argument based around EC* self-defeating.

When it comes to avoiding self-defeat, we have drawn three lessons. First, if a debunker does not provide some reason to think that EC* is e-connected, any debunking argument employing it is *prima facie* self-defeating. A debunker relying on it must therefore adopt some metaepistemological account that shows it to be e-connected.

Second, a non-conditional debunker who defends some such metaepistemological account will be faced with the question of whether that account also shields belief in moral facts. If it does, the debunking argument becomes toothless. If it does not, the debunker is forced to deny moral-epistemic parity and provide the requisite metanormative account to back that up.

Third, a conditional debunker can avoid self-defeat, and avoid having to deny moral-epistemic parity, by supplementing their debunking argument with a unified, positive metanormative account of the moral and epistemic domains. In this way, a conditional debunker shifts their explanatory burdens over to their metanormative account.

The main takeaway from the chapter is that debunking arguments relying on explanatory constraints require a debunker to stake out a position on relatively unexplored and deeply contested issues in metanormativity and metaepistemology.

8 Looking Back, Moving Forward: Debunking without Self-Defeat

8.1 Introduction

In this concluding chapter, we will return to the central issues set out in Chapter 1 and consider how contemporary debunking arguments fare with respect to them (§8.2). We will then apply and generalize the lessons from previous chapters and consider the prospects for future formulations of global moral debunking arguments (§8.3). In particular, we will consider what challenges conditional and non-conditional debunking arguments face, and how they could mitigate the costs of running a successful debunking argument.

For non-conditional debunking arguments, I will suggest two conditions of adequacy that such arguments will need to satisfy in order to avoid succumbing to either self-defeat or toothlessness. For conditional debunking arguments, I will suggest that they must begin to take their metaepistemological commitments seriously and that it is no longer possible to serve up debunking arguments that only consider the moral domain. We then move on to consider two other important lessons from the preceding chapters involving the relative unimportance of normativity and the importance of the a priori. I end the chapter with a final conclusion (§8.4).

8.2 Central Issues, Revisited

In Chapter 1, I listed four central challenges that have plagued the project of constructing a successful global moral debunking argument. These issues concern finding the right epistemological principles for debunking, avoiding overgeneralization and self-defeat, as well as the importance of a debunker's metaepistemological commitments. Having now looked at a broad range of contemporary debunking arguments in the previous chapters, we are in a position to evaluate where debunking arguments stand with respect to these issues.

8.2.1 Epistemological Principles

In Part I, we saw that early, paradigmatic debunking arguments often lacked an explicit explanation of how, exactly, genealogical information could suffice to power a global moral debunking argument. Instead, such arguments relied on evocative analogies or on an appeal to various vague or underspecified epistemological notions such as moral belief being 'off-track', 'insensitive to evidence', 'coincidental', or bearing an 'independence relation' to the moral facts.

In the chapters that followed, we saw examples of how second-generation debunking arguments pick up one or more aspects of such early debunking arguments—explanatory relevance, modal constraints, unexplained reliability—and sharpen them into more precise principles. Debunkers following Enoch's cue, have zeroed in on the aspects of Street's argument that deal with the coincidental nature of the reliability of moral beliefs.

Debunkers who have continued the tradition from Joyce of arguing that moral beliefs are insensitive to evidence have adopted modal conditions that are either meant to be necessary conditions for knowledge, or sufficient conditions for epistemic defeat. These are taken to be the operative mechanism that allows genealogical information to globally undermine moral belief. This project is, as we saw, severely underdeveloped. This is in large part because the epistemology of justified belief in necessary propositions and a priori knowledge is relatively underexplored. As these fields develop, we are likely to see new debunking arguments applying more sophisticated modal conditions.

Lastly, debunkers like Korman and Locke, have picked up the explanationist bent of early debunking arguments, such as Harman's, and sharpened it into epistemological principles that explain how genealogical information can underwrite the uniform undermining of moral beliefs, without invoking coincidental reliability or modal conditions. When targeting the moral domain in isolation, such arguments might seem promising, but when considered against the backdrop of generalization and self-undermining they need to be accompanied by extensive additional theoretical commitments which makes them much more costly than has previously been recognized.

Debunking arguments have therefore been developed in different directions, which in turn result in such arguments being subject to different objections. What is true of all these arguments is that they attempt to seek out some general epistemological principle—e.g., UNEXPLAINED RELIABILITY, SAFETY, EC*—that explains how debunking arguments debunk. While we have seen that it is challenging to defend such general first-order epistemological principles, it would seem implausible to retreat to the claim that it would suffice to rely on an illustrative analogy or merely consider it on a case-by-case basis. The implausibility of the latter approach is seen by the fact that a debunking argument that seeks to debunk a belief that p on the grounds that this belief has some epistemically defective feature, F, which renders it epistemically defective, should be expected to debunk *all* beliefs possessing F.

If we accept that debunking arguments do need to subscribe to some such general first-order epistemological principle, what would this mean for the prospects of debunking morality? The principles we have considered have, almost without exception, been highly contested. This speaks against a strain of optimism regarding debunking arguments, which has seen them as a particularly dialectally effective type of argument.[1] Consider the reliability challenge, discussed in Chapters 5 and 6, which was supposed to be able to grant the moral non-naturalist pretty much all they could ask for—defeasible justification for their moral beliefs, the metaphysical necessity of moral facts, and that moral facts are irreducible, stance-independent, and causally inert—but still result in the non-naturalist not being able to explain the reliability of their moral beliefs.

Such an argument would be guaranteed to get traction with its intended target, the non-naturalist, since it exclusively employs premises that a non-naturalist antecedently accepts. The necessity of introducing highly contested general epistemological principles as premises more or less precludes the possibility of debunking arguments gaining this kind of dialectical traction. At least unless the target of a debunking argument happens to be antecedently committed to the relevant epistemological principle employed by the debunking argument in question.

It might be thought unsurprising that controversial philosophical arguments will, in the end, need to appeal to highly contested premises. It will nonetheless entail that the debate over the prospects of debunking arguments will move away from discussions of the moral domain, and to a debate over general epistemology and metaepistemology. This move has already begun to take place in the literature, which in turn has caused the debate over debunking arguments to splinter into a debate over the correct first-order epistemological principles, both as a general epistemological project, and for the purposes of debunking morality.

[1] Enoch (2011, 158).

Defending an epistemological principle as the operative principle in a debunking argument generates certain challenges for debunkers, which have rarely been acknowledged. One such task is to consider not only how a debunking argument fares with respect to its target domain, but whether the principle employed causes the argument to generalize beyond it.

8.2.2 (Over)generalization

Many, if not most, debunking arguments have been presented as targeting either one or a few domains, such as morality or religion. Most formulations of such arguments have paid little attention to whether their argument would, in fact, apply equally to other domains. It is not uncommon for authors to simply state, often in a footnote, that they will not be concerned with whether a given debunking argument generalizes, or even if the argument faces the threat of self-defeat. Despite this, we have seen that most of the debunking arguments we have considered risk generalizing beyond their intended domain, often to the detriment of the plausibility of the argument in question.

Whether or not the generalization of an argument to a different domain is a feature or a bug is to some extent up to the predilections of the debunker. Even so, robbing certain beliefs of their positive epistemic status should be anathema to most. While we saw that some principles, such as EC*, risk undermining belief in empirical propositions, most principles saw a greater risk of generalizing to non-empirical domains. This is, I believe, because of the deep, structural similarities across domains that are often considered non-empirical, such as morality, metaphysics, mathematics, modality, logic, and epistemology.

If a debunker wants to target moral non-naturalist realism for its problematic epistemology, they had better be prepared to either target structurally similar views about other non-empirical domains or propose a conception of such domains on which they are exempted, for instance, because they are argued to in fact be empirical domains. This is an enormous task, as the epistemology of these domains is notoriously contested.

I believe no debunking argument can be properly evaluated without having a clear picture of the extent to which it generalizes. This is true, not least, because some forms of generalization are fatal.

8.2.3 Self-defeat

Whether or not one finds generalization to other domains to be a bug or a feature of a debunking argument, one cannot allow it to generalize in a way that would target the debunkers belief in the epistemological principles employed by the argument itself. We explored this issue in depth for explanatory constraints in Chapter 7.

It is worthwhile to consider to what extent the findings of Chapter 7 generalize to other, non-explanationist debunking arguments. Let us therefore briefly consider whether it applies to UNEXPLAINED RELIABILITY, discussed at length in Chapter 5. Recall,

> UNEXPLAINED RELIABILITY
> If the reliability of S' belief that p cannot be explained, S' belief that p is undermined.

Imagine for a moment that the reliability challenge was successful in showing that, on a non-naturalist conception of moral facts, our moral beliefs are uniformly undermined. The debunker launching that argument would be relying on the epistemological principle UNEXPLAINED RELIABILITY. A non-naturalist would be well within their rights to question whether the principle is self-defeating. The question therefore becomes whether the reliability of epistemic beliefs, like a belief in UNEXPLAINED RELIABILITY, can be explained.

Recall that the difficulty involved in explaining the reliability of moral beliefs arose because there seemed to be a distinct lack of plausible explanatory models for how to explain the correlation between our moral beliefs and moral truths—the MINIMAL CORRELATION. For a non-naturalist, this was because non-natural moral facts are stance-independent, causally inert, and irreducible.

What about epistemic facts? Assume first that such facts are stance-independent, irreducible, and causally inert.[2] This would eliminate the possibility of our attitudes explaining the moral facts by way of constitutive explanations. A principle such as UNEXPLAINED RELIABILITY might not, on the face of it, seem like the type of fact to be involved in causal relations either. Neither does such epistemic facts look like an obvious candidate for being reduced to, identified with, wholly constituted by, or fully grounded in some natural fact. Lastly, the outlook for some form of epistemic rationalism that could secure a

[2] See Boghossian (2006) and Cuneo (2007) for a defense of such claims.

direct rational explanatory connection between epistemic facts and epistemic beliefs should be no better than for moral beliefs.

Taken together, this provides a *prima facie* case that there is no *direct* explanatory connection between our epistemic beliefs and the epistemic facts. This rules out the first, direct, model for explaining the correlation.

Could our epistemic reliability be explained by a third factor? In Chapter 5, we saw that a genuine third-factor explanation is not available to the moral non-naturalist. That was because such theorists deny that any natural facts (or non-natural facts) participate in the explanation of fundamental moral facts. We are assuming epistemic facts to be stance-independent, causally inert, and irreducible. This would make it plausible that the fundamental epistemic facts are ungrounded as well.[3]

If this is indeed the intuitive way to construe epistemological facts like UNEXPLAINED RELIABILITY, then a third-factor explanation could not explain the reliability of epistemic beliefs. This therefore rules out the second, indirect, model of explaining the correlation.

Could the reliability of our epistemic beliefs be explained by an appeal to a cosmic coincidence? Insofar as the debunker will want to block this option in the moral case, they will presumably want to stay away from themselves embracing it in the epistemic case.

What about endorsing full-blooded Platonism about the epistemic domain? This is unlikely to be attractive since such a view would suffer the same fate as when applied to the moral domain. There could then not be any unique, non-theory relative answer to questions about how we should conduct our doxastic practice.[4]

If UNEXPLAINED RELIABILITY is true, and if the debunker is correct that we cannot explain the reliability of our moral beliefs, then moral beliefs are thereby uniformly undermined. If epistemic facts share the problematic features with moral facts, then we are equally unable to explain the reliability of our epistemic beliefs. As UNEXPLAINED RELIABILITY is an epistemic fact, any (non-conditional) debunking argument relying on it would therefore, at least *prima facie*, target our belief in the principle itself and be self-defeating.

In the above, we assumed epistemic facts to be stance-independent, irreducible, and casually inert. Surely, a debunker need not commit to this. However, at this point, we can generalize the central lesson for non-conditional

[3] This assumes that the structure of explanation of epistemic facts is similar to the structure of moral explanations set out in 3.6.

[4] See §5.3.2 for discussion of this in the moral case. Interestingly, such a conclusion might perhaps be more acceptable for the epistemic domain than the moral domain.

debunking arguments set out in the previous chapter. In order to avoid such *prima facie* self-defeat, a debunker must do three things:

(a) explain why epistemic beliefs are not subject to the epistemic defect attributed to the target beliefs, and,

(b) show that the beliefs targeted by their debunking argument *do* possess this defect, and,

(c) show that their strategy for establishing (a) does not generalize so as to also shield the beliefs targeted.

The same costs that we saw arising for explanationist debunking arguments will therefore arise for other debunking arguments as well, such as for the reliability challenge. For this reason, the threat of *prima facie* self-defeat is a general feature of most debunking arguments, forcing debunkers to clarify (a)–(c) in order to avoid it.

I now want to diagnose the reason why many popular global moral debunking arguments are *prima facie* self-defeating in this way. This is because, if moral and epistemic facts are relevantly similar, such debunking arguments will attribute to a belief in its own operative epistemological principle *exactly* the flaw that it attributes to the moral beliefs it targets. To see why this is the case for many, if not most, global moral debunking arguments, consider the features of moral facts and beliefs that such debunking arguments have claimed, in some way, to undermine moral beliefs:

(1) *Explanatory dispensability:* Moral facts are not indispensable for explaining any of our moral (or non-moral) judgments[5]

(2) *Unexplained Reliability:* The reliability of our moral beliefs would have to be a coincidence[6]

(3) *Insensitivity:* We would still have our moral beliefs, even if they were false[7]

[5] Harman (1977); Joyce (2006).
[6] Street (2006); Enoch (2011).
[7] Joyce (2001, chap. 7); Braddock (2017).

(4) *Lack of safety*: We could have easily had false moral beliefs[8]

(5) *Explanatory constraints*: Our moral beliefs fail to satisfy some explanatory constraint which is sufficient for epistemic defeat[9]

(6) *Disagreement:* Moral facts are subject to deep, widespread disagreement[10]

These are the usual suspects when it comes to explaining why moral beliefs are epistemically defective, and often constitute, independently or in some combination, the grounds upon which debunking arguments are constructed. I have argued that, at least as far as non-naturalists are concerned, they are vulnerable to such debunking arguments in virtue of ascribing to moral facts the following three features:

(i) *Stance-independence:* Fundamental moral facts do not obtain in virtue of the attitudes, practices, or customs of human beings.

(ii) *Causal inertness:* Moral facts lack causal powers.

(iii) *Irreducibility:* Moral facts are not reducible (without remainder) to, identical with, wholly constituted by, or fully grounded in natural facts.

What I now want to suggest is that any debunking argument that attempts to leverage any of the features in (1)–(6), or any relevantly similar features, when constructing a debunking argument will run into *prima facie* self-defeat. The reason for this, which should have become apparent by now, is that on a metaepistemic conception that involves the following three claims, epistemic beliefs will be equally vulnerable to the different types of debunking challenges stemming from (1)–(6):

(i*) *Stance-independence:* Fundamental epistemic facts do not obtain in virtue of the attitudes, practices, or customs of human beings.

(ii*) *Causal inertness:* Epistemic facts lack causal powers.

[8] Cf. Srinivasan (2015); Clarke-Doane (2020).
[9] Korman and Locke (2020); Lutz (2020).
[10] Tolhurst (1987).

(iii*) *Irreducibility:* Epistemic facts are not reducible (without remainder) to, identical with, wholly constituted by, or fully grounded in natural facts

If (1)–(6) are grounds for taking our moral beliefs to be undermined, then they are equally good grounds for taking our epistemic beliefs to be undermined, given an endorsement of (i*)–(iii*).[11] I have already argued for this with respect to (1)–(5), and I will now briefly say something about (6)—debunking arguments that take some form of disagreement to undermine moral beliefs.

Many have argued that widespread moral disagreement, at least when properly qualified, uniformly undermines moral beliefs.[12] Such qualifications might for instance include that the disagreement must be between epistemic peers. The formulation of any debunking argument based on widespread moral disagreement will rely on epistemological principles. Those epistemic premises will, either explicitly or implicitly, rely on some principle which rules widespread, deep disagreement—or some such—an epistemic defect sufficient for undermining moral beliefs.

A cursory glance at contemporary epistemology, as well as the discussion in previous chapters, should be sufficient evidence to show that there will be no less, and perhaps even more, widespread disagreement about epistemological principles and epistemic facts. This will likely hold true both when it comes to disagreement about principles and particular cases, as well as between peers who are specialists and peers who are not. While settling the issue would require extensive empirical investigations, it is hardly controversial that there is widespread disagreement among epistemologists about principles.

Similarly, it can hardly be doubted that non-specialist peer disagreement about, say, what to believe, is ubiquitous. Perhaps even more so than in the moral case. In all likelihood, this is the case for whatever particular principle the debunker relies on as well. If deep, widespread disagreement, with some additional qualifications, is sufficient to undermine moral belief, it would seem equally successful, at least *prima facie*, in undermining epistemic beliefs, such as the premises of the debunking argument.[13]

[11] Cuneo (2007) defends a version of this claim, as well as a view of epistemic facts that is very similar to the conjunction of (i*)–(iii*). Cuneo uses ‘irreducible’ in a sense that might render his view, at least in principle, compatible with reductive views (cf. Heathwood 2009, 84). See also Boghossian (2006) for a defense of such a view of epistemic facts, developed through an argument against epistemic relativism and contextualism.

[12] For an overview of such debates, see Gowans (2000) and Rowland (2021).

[13] Sampson (2019) develops this type of self-defeat challenge facing arguments from disagreement.

In the case of debunking arguments based on disagreement, the threat of self-defeat is less obviously dependent on endorsing a conception of epistemic facts which includes (i*)–(iii*). Because of this, such arguments might seem to have an *even harder* time fending off the self-defeat challenge, as debunkers endorsing them cannot merely appeal to some metaepistemological strategy that denies one or more of those commitments.[14] So while (i*)–(iii*) are what tends to generate the threat of *prima facie* self-defeat, this is not always so, as the case of disagreement shows.

We will return to how different debunking arguments can avoid self-defeat in §8.3. Before considering that issue, lets us consider the underappreciated importance of metaepistemological commitments for debunking arguments.

8.2.4 Metaepistemological principles

Let me now say something more detailed about the conception of epistemic facts that underlies the threat of *prima facie* self-defeat. Above, I set out the features (i*)–(iii*) that one might take epistemic facts to possess. These features single out a non-reductive realist metaepistemological account of epistemic facts (hereafter *epistemic non-reductionism*).[15] Such an account is strikingly similar to a moral non-naturalist's account of moral facts, and the two views' parallel each other to an important degree.

One similarity is that epistemic non-reductionism is incompatible with metaepistemological views that hold epistemic facts to be reducible (without remainder) to, identical with, wholly constituted by, or fully grounded in natural facts. Similarly, it is incompatible with views that hold epistemic facts to be capable of participating in causal explanations, or that fundamental epistemic facts are in some way constituted by or constructed from our attitudes or practical standpoint. In other words, epistemic non-reductionism is incompatible with *epistemic reductionism*, which holds that epistemic facts can be reduced (without reminder) to, identified with, wholly constituted by, or fully grounded in some set of natural facts.[16] Similarly, it is incompatible with *ep-*

[14] It could appeal to a different metaepistemological strategy than those I have mentioned so far. One suggestion—involving self-exempting principles—is discussed by Elga (2010).

[15] Boghossian (2006); Cuneo (2007). In addition to (i*)–(iii*), the view would take on a number of claims analogous to those outlined in §1.4.

[16] Heathwood (2009, 85; 2018).

istemic constructivism and *epistemic institutionalism,* which holds that the epistemic facts, in some way, depend constitutively on our attitudes, our practical or agential nature, or on social conventions.[17]

I suggested above that a debunking argument based around (1)–(6), or on relevantly similar grounds, faces *prima facie* self-defeat. This threat of self-defeat makes it incumbent on anyone presenting a debunking argument to show how it manages to avoid it. In order to avoid such self-defeat, a debunker will have to show that whatever their conception of epistemic facts is, it does not render their epistemic beliefs vulnerable to debunking of the same kind. This means that debunkers must be explicit about their commitments to some metaepistemological strategy that avoids this outcome.

I believe that no global moral debunking argument can be properly evaluated without having a clear picture of to what extent it is self-defeating. Since, as I have claimed, debunking arguments face *prima facie* self-defeat, this means that no debunking argument can properly be evaluated without spelling out its metaepistemological commitments.

In the next section, I will argue that this imposes conditions of adequacy on any such debunking argument, which few extant arguments have managed to satisfy.

8.3 Future Directions

8.3.1 Two Conditions of Adequacy

Debunking arguments rely on first-order epistemological principles. A debunker therefore cannot defend a moral debunking argument that generalizes in a way that ends up undermining their belief in their operative principle. In order to avoid self-defeat, a debunker must therefore show that their belief in their operative epistemological principle is not itself targeted by their argument. Call this task NO SELF-DEFEAT.

[17] Constructivism: Warenski (2021, sec. 4); Street (2009). Conventionalism: Côté-Bouchard and Littlejohn (2018, 159–60), Woods (2018), Cowie (2019); Mantel (2019); Maguire and Woods (2020).

> NO SELF-DEFEAT
> Whatever feature a debunking argument's epistemological principle rules an epistemic defect must not attach to the debunker's belief in that very epistemological principle.

While such a condition of adequacy might seem blindingly obvious, we have seen that many, if not most, debunking arguments have failed to convincingly satisfy it. What is needed to avoid it, is for a debunker to lay out their metaepistemological commitments in a way that shows them not to be subject to self-defeat.

In Chapter 7, we saw a number of possible metaepistemological strategies that are capable of allowing a debunker to satisfy NO SELF-DEFEAT. These include embracing certain forms of epistemic rationalism, as well as variants of reductive naturalism, constructivism, and expressivism. A debunker might also be able to defend the claim that epistemic facts are explanatorily indispensable, that they are inductive principles that do not apply to themselves, or that they are, in some sense, institutional facts.[18]

Committing to some such metaepistemological strategy is not sufficient to discharge the self-defat challenge as I have posed it. This is because many of the strategies that would allow a debunker to satisfy NO SELF-DEFEAT would risk generalizing so as to also protect the beliefs that are the intended target of the debunking argument. We saw that this is less of an issue for conditional debunking arguments, so I set them aside for now.

Insofar as a (non-conditional) debunker wants to commit to any metaepistemological strategy, they will, in addition to NO SELF-DEFEAT, need to show that their preferred strategy does not work equally well for whatever beliefs their debunking argument is intended to target. Call this second task DISPARITY.

> DISPARITY
> Whatever metaepistemological strategy a debunker employs to shield their belief in the relevant epistemological principle(s) from self-defeat cannot also shield the beliefs targeted by their debunking argument from defeat.

[18] For indispensability, see Cowie (2019, sec. 10.3). For "unmodest inductive methods", see Lewis (1971). Elga (2010) apply such a strategy to block self-defeat worries about certain principles of peer disagreement. On epistemic facts being institutional, see Côté-Bouchard and Littlejohn (2018, 159–60), Woods (2018), Cowie (2019), Mantel (2019), and Maguire and Woods (2020).

If a debunker manages to protect their belief in the operative epistemological principle, but fails to block the application of this strategy to the beliefs targeted by their argument, their argument is rendered toothless. However, in satisfying DISPARITY, the debunker will have to deny widely accepted claims about moral-epistemic parity.

I believe most (non-conditional) debunkers have failed to appreciate these concerns, and that this has made them overestimate the plausibility of their arguments. Many debunkers have singled out their target domain, often morality, and presented arguments against it without an eye toward the larger implication and potential reach of their arguments. This is likely because they have failed to realize the full generality of the epistemological principles they have utilized, implicitly or explicitly.

I submit that the project of investigating the extent to which non-conditional debunkers can satisfy both NO SELF-DEFEAT and DISPARITY is a crucial one. It is, however, a major undertaking. It will involve mapping out the territory of metaepistemology in a way comparable to what has been achieved in metaethics. It will also involve evaluating the plausibility of the theories that make up the metaepistemological space in their own right. With that evaluation in place, it will be necessary to consider whether any plausible metaepistemological strategy is able to satisfy the two tasks set out above.

While the recent literature on debunking arguments has started to get explicit and systematic when it comes to mapping out their first-order epistemological commitments, their metaepistemological commitments still tend to be shrouded in obscurity. Bringing clarity to that issue will be the central task for debunkers going forward.

8.3.2 Conditional Debunking Arguments

Conditional debunking arguments, when supplemented with a positive, unified metanormative proposal, face an entirely different challenge than the non-conditional debunker. Take a non-conditional debunker who targets moral non-naturalism. As long as the conditional debunker opts for a unified metaepistemological view that is sufficiently different from epistemic non-reductionism, they will manage to satisfy NO SELF-DEFEAT. And they will be able to do so without having to worry about establishing DISPARITY, as the non-naturalist cannot opt for any alternative view in the moral domain.

Such success comes at a price. To achieve it, a debunker will need to defend a unified, metanormative view encompassing, at the very least, both epistemic and moral facts. To avoid problems stemming from generalization, they might also need to extend that theory to other domains traditionally considered a priori, such as metaphysics, mathematics, modality, and logic. This, to be sure, is no small task. Admirably, Street has pursued exactly this tactic. When evaluating her debunking argument, however, her positive constructivist proposal is often completely glossed over. This, as we have seen, would be a mistake.

8.3.3 The Irrelevance of Normativity

I have argued that many aspects of the epistemology of debunking arguments have been under-appreciated. I now want to draw attention to an aspect of such arguments which I believe has been overestimated. Note that no important part of any debunking augment that we have discussed so far has relied on the claim that epistemic facts (or moral facts) are irreducibly *normative*, or really, normative at all.

In contrast, Christos Kyriacou has argued that debunking arguments, in general, are self-defeating because they target normative facts and properties (e.g., moral facts and properties), while themselves relying on just such normative facts and properties (e.g., epistemic facts and properties).[19] He claims that debunking arguments charge various belief-forming processes with being unreliable, in particular those processes that generate belief in normative facts and properties. Debunking arguments themselves rely on premises involving epistemic facts, and properties, which are, arguably, normative. A debunking argument will therefore target beliefs in the type of facts, and properties that it itself presupposes, or which figures as its premises, and will therefore be self-defeating.

Kyriacou thinks the threat of self-defeat arises from targeting *normative* or *irreducibly normative* beliefs. I believe this is something of a red herring. Consider the debunking arguments we have looked at so far. Those arguments have relied on the assumption that epistemic facts are irreducibly *epistemic* (and that moral facts are irreducibly *moral*). Of course, epistemic and moral facts could be held to *be* normative, but that is not the feature that the debunking arguments target. This can be seen in the way that debunking arguments generalize to also target mathematical, logical, or modal facts, when these facts are taken to not reduce to natural facts. As far as we have been concerned, normativity has been an entirely irrelevant aspect of such facts and propterties.

[19] Kyriacou (2016).

We have seen that the central issues for debunking arguments have revolved around explanatory relevance, explanations of reliability, stance-independence, causal inertness, and irreducibility. While normative domains usually face challenges based on these features, there seems to be no special connection between them and normativity as such.

We have found that the issues that arise in connection with debunking arguments are often shared with non-normative domains. Similarly, as evidenced by the argument in Chapter 7, there need be nothing essentially normative about the self-defeat challenge. Korman and Locke's explanatory constraint is faced with self-defeat on grounds that are unrelated to its normativity. This suggests that it is not the purported irreducible normativity of epistemic facts that gives rise to the self-defeat challenge.

Whatever worries one might have about irreducible normativity, they are unlikely to undergird debunking arguments themselves or the threat of self-defat that faces them.

8.3.4 The Importance of the A Priori

Debunking arguments face a puzzling predicament. Despite being intuitively compelling and having been widely discussed, they seem beleaguered by deep-running difficulties stemming from very general epistemological and metaepistemological considerations. How could it be that any global moral debunking argument could be so theoretically costly and difficult to successfully pull off? I believe that, rather than having to do with normativity, this stems from the unity of a priori domains.

I have argued that what gives rise to the threat of self-defeat are, at heart, worries concerning stance-independence, causal inertness, and irreducibility. Moral facts are often taken to have some, or all, of these features. To the extent that they do, moral belief is subject to several puzzles, including not least how we are to understand the relation between our beliefs and the relevant a priori justified propositions. This raises questions such as: If our beliefs have a causal history that does not at any point include moral facts, would we believe what we do no matter what the facts were? What explains our success in forming moral beliefs? Could we easily have had false moral beliefs? And are our beliefs *in any way* connected to the facts they are about?

There is seemingly nothing distinctly moral about these puzzles. They would seem pressing for all traditionally a priori domains. On this construal, the problems stem from general questions concerning a priori knowledge.

Some domains that are typically thought to involve such knowledge include metaphysics, epistemology, logic, modality, and mathematics. To the extent that the features that spark debunking arguments have to do with the general epistemic features of a priori knowledge, it is to be expected that all of these domains will face similar challenges. And they do. The question is therefore whether there is something that differentiates these domains sufficiently from each other such that one can run a debunking argument against only one or two, without equally targeting all a priori domains. Such questions will require patient, in-depth investigation and deep knowledge of the relevant fields.[20]

Insofar as the above line of reasoning is correct, there should be nothing particularly debunkable about morality, or even normative domains more generally. Rather such worries would likely extend to all domains where we, at least *prima facie*, attribute stance-independence, causal inertness, and irreducibility to the relevant set of facts and/or entities.

Empirical ways of knowing involve being in causal contact with whatever it is we come to know, where the relevant facts help explain our beliefs by way of familiar chains of causal relata. Such causal chains, in favorable circumstances, lead to sensitive, safe, and explanatorily well-connected beliefs. While still raising numerous questions, we do seem to have something of a grasp of how such empirical knowledge is possible.

We usually take ourselves to also know things for which it is not plausible to think that we have come to know them on the basis of causal contact and empirical evidence. How could we know—in the empirical sense—whether maximizing utility makes an action right, whether sensitivity is required for knowledge, or whether there are fewer and fewer prime numbers the larger they get?

If we are to explain our purported knowledge in these domains, we seem forced to appeal to a qualitatively different way of knowing. Accounts of non-empirical knowledge, unfortunately, are far less developed than their empirical counterpart. Given this situation, I want to outline a few possible methodological stances one could take in response to this state of affairs vis-à-vis debunking arguments. I think these stances represent the three main camps we have covered in the preceding chapters—the non-naturalist, as well as the conditional, and the non-conditional debunker.

Insofar as we think our knowledge of non-empirical domains is fairly secure, as the non-naturalist is wont to do, we might simply content ourselves

[20] An exemplary instance of such an in-depth, comparative investigation, is Clarke-Doane's (2020) exploration of the parallels and disparities between morality and mathematics.

that we will come up with some acceptable account of non-empirical knowledge in the future.[21] Such a conservative stance would allow us to take large swaths of our current beliefs and knowledge at face value. This would be akin to Newton's acceptance of gravity being an 'occult force' operating at a distance—we might not like it, but the "data" requires it.

It might be that we are therefore justified in maintaining beliefs in a priori propositions, as well as taking ourselves to possess a priori knowledge, despite the mechanisms behind our epistemic success in this domain being unknown to us. This could be so, for instance, if discounting the evidence would be, on balance, even less plausible.

A second stance would be to reinterpret those domains that fall outside of our empirical knowledge—mathematics, logic, modality, morality, epistemology, metaphysics—along the lines of traditional empirical domains. This is, of course, the project of "naturalization." It involves reformulating our conceptions of these domains in ways that render them instances of more or less camouflaged empirical knowledge.[22] Why not, then, attempt to refigure all instances of non-empirical knowledge—mathematical, logical, modal, metaphysical, moral, and epistemic—into the mold of empirical knowledge?

This is what most conditional debunkers could be understood as attempting. Such undertakings are formidable, and it can be difficult to see how they could possibly succeed. In their favor, they would allow us to avoid accepting that epistemic processes we do not understand underlie much of our knowledge. If certain pockets of purported non-empirical knowledge cannot be reinterpreted according to an empirical model, then so much the worse for it.

For such domains where naturalization fails, a third approach could be taken. Concluding that the obstacles to a reconceptualization of one or more of these domains in the framework of empirical knowledge are insurmountable, this approach will instead undertake the project of giving up the domains. This has been the role of non-conditional debunking arguments. If we do not know how we could non-empirically know that p—and no naturalizing project seems plausible—why not withhold belief about whether p?

While this might seem perfectly reasonable, we have seen repeatedly that the project of explaining why it would be rational to give up various forms of non-empirical knowledge *relies on* principles that are standardly taken to fall under the banner of the domains that are being purged. It therefore becomes

[21] Huemer (2016, 1986); FitzPatrick (2015, 894).

[22] For an extreme defense of this view, which abandons a priori knowledge altogether, see Devitt (2005; 2011).

necessary, even for the skeptic, to first reinterpret at least parts of the epistemic domain into an acceptable substitute. This is needed to undergird the project of eliminating whatever purported non-empirical knowledge the debunker chooses to target.

As I have tried to show in these pages, there is something puzzling about this last strategy. Debunkers tend to target some non-empirical domain. Doing so requires them to take on the naturalizing project with respect to *one* domain that is traditionally considered non-empirical, epistemology. However, once they begin engaging in that project, one could think that what is good for the goose is good for the gander. And so the strategy for naturalizing one domain could be applied to other non-empirical domains as well. And then there is not much left to debunk. This is a tension we have seen materialize for the non-conditional debunker in the form of the need for establishing a disparity between the epistemic domain a debunker relies on and the moral domain they seek to target.

In their own way—and understood as large-scale epistemological projects far from being completed—all of these stances are perfectly sensible. This is perhaps why, when they face off against each other, they often result in a perfectly legitimate stalemate.

8.4 Conclusion

My goal in this thesis has not been to establish that it would be impossible for a debunker to avoid self-defeat, nor that it is impossible to run a successful debunking argument without incurring prohibitive theoretical costs. Rather, it has been to show that developing a successful global moral debunking argument is going to be, without exception, extremely theoretically costly in terms of the required concomitant commitments. I have shown what these costs consist in and suggested some ways in which a debunker could attempt to get out of their explanatory debt as cheaply as possible.

Swedish Summary

Under de senaste decennierna har intresset för en gammal filosofisk fråga ökat explosionsartat: När borde vi avfärda uppfattningar på grund av var de kommer ifrån? Närmare bestämt, under vilka omständigheter bör vi ge upp en uppfattning därför att vi har blivit medvetna om att den har ett tvivelaktigt ursprung? I stora drag är det denna fråga avhandlingen försöker besvara.

Under lång tid har man ansett att det är ett misstag att tro att en uppfattnings ursprung överhuvudtaget kan påverka dess trovärdighet. Detta kallas ibland för *det genetiska argumentationsfelet*. På senare tid har det dock blivit alltmer accepterat att hävda att ursprunget till våra uppfattningar kan ge oss skäl att avfärda dem. Denna ståndpunkt har drivits i diskussioner om så kallade ”undergrävande förklaringar”. Dessa är förklaringar som ska visa att vi inte längre har goda skäl att hålla fast vid en uppfattning.

Föreställ dig till exempel följande: Tidigt en morgon är du övertygad om att någon har brutit sig in i ditt trädgårdsskjul och stulit din favoritblomkruka. Efter att ha druckit ditt morgonkaffe går det upp för dig att du bara hade en ovanligt livlig dröm den natten om en tjuv som stal din älskade kruka från skjulet. När du inser att övertygelsen om den stulna krukan kom från en dröm, avfärdar du den med ett skratt.

Här finns byggstenarna till en undergrävande förklaring. Här har vi nämligen information om en uppfattnings kausala eller historiska ursprung—alltså om dess *genealogi*—som upphäver uppfattningens trovärdighet. Du hade en uppfattning—att någon stal din blomkruka—och du hade ingen anledning att misstro den. Men sedan inser du att uppfattningen kom från en notoriskt opålitlig källa: drömmande. Att få reda på att källan till din uppfattning är denna opålitliga process är tillräckligt för att neutralisera dess *epistemiska* trovärdighet—tillräckligt för att göra det irrationellt, omotiverat eller på annat sätt olämpligt för dig att fortsätta att hålla fast vid den. Observera att detta är fallet även om det, utan att du vet om det och av helt oberoende skäl, ändå vore sant att någon stal din blomkruka.

Ett sådant scenario visar att vi ibland borde ge upp en uppfattning, utan att vi direkt har fått skäl att tro att uppfattningen är falsk. Allt vi har fått veta är att uppfattningen kom till genom en opålitlig process. Kanske verkar detta smärtsamt uppenbart—självklart kan kunskap om en uppfattnings ursprung undergräva dess trovärdighet! Så har det också verkat för många filosofer. Argument som åberopar våra trosföreställningars ursprung för att avfärda dem har blivit alltmer framträdande inom många filosofiska underdiscipliner. Många har argumenterat för att vi kan hitta denna typ av undergrävande förklaringar när det gäller våra moraliska övertygelser, till exempel genom att hänvisa till empiriska påståenden om ursprunget till vår moraliska kognitiva arkitektur eller evolutionära influenser på våra trosuppfattningar. Denna typ av argument kallas för genealogiska debunking argument (härefter bara *debunking argument*), och är det centrala fokuset i avhandlingen.

Många hävdar på denna basis att det inte är någon tillfällighet att sådana saker som skademinimering, omhändertagande av våra barn och socialt samarbete är av största vikt i våra moraliska liv. Detta är precis vad vi kan förvänta oss om vi är varelser som tillhör en släkt vars förfäders moraliska attityder påverkats avsevärt av evolutionära urvalstryck. Enligt detta tankesätt har sådan påverkan lett till att våra förfäder har haft positiva attityder till sådant som är nära kopplat till ökad reproduktiv framgång, som till exempel överlevnad, barns välbefinnande och samarbete inom grupper.

Många har tänkt att detta borde få oss att omvärdera våra moraliska uppfattningar. Ett skäl till detta, enligt många, är att vi skulle kunna se den nuvarande förekomsten av moraliska övertygelser, även om dessa övertygelser var falska. Detta eftersom urvalstrycket gynnade dem som hade en positiv inställning till beteenden som främjar överlevnad och samarbete, enbart därför att denna typ attityder ledde till ökad reproduktiv framgång.

Den här typen av resonemang ger upphov till en rad kritiska frågor. Har vi våra moraliska uppfattningar endast eftersom vi är djur av en viss sort? Skulle en sådan evolutionär ursprungshistoria för våra moraliska uppfattningar, när den är klarlagd och givits empiriskt stöd, tvinga oss att ge upp dessa uppfattningar? Och skulle vi annars riskera att vara irrationella? På senare tid har många svarat ”ja” på dessa frågor. Om så är fallet, så har vi en undergrävande förklaring av våra moraliska uppfattningar.

Att acceptera en undergrävande förklaring av våra moraliska uppfattningar ger i sin tur upphov till en rad frågor. Varför ska vi tro att evolutionär påverkning är den enda form av tvivelaktigt ursprung som kan undergräva våra uppfattningar? Det finns säkert otaliga andra dolda influenser som på samma sätt påverkar vad vi tror—kulturell och social tillhörighet, historisk period, kön, religion, uppfostran och så vidare. Har inflytande från sådana faktorer på samma sätt makt att undergräva trovärdigheten hos våra uppfattningar? Är det tillräckligt att visa att en uppfattning har några av dessa influenser bland sina kausala eller historiska rötter för att tvinga oss att ge upp den? Dessa frågor utgör fundamentet för problemställningen avhandlingen försöker besvara: När ger ursprunget till en uppfattning oss skäl att ge upp den?

Avhandlingsprojektet syftar till att utvärdera utsikterna och utmaningarna för argument som kommer till slutsatsen att vi inte har goda skäl för våra moraliska uppfattningar på grund av någon form av undergrävande förklaring som visar detta. Målet är därför att etablera vad som skulle krävas för att konstruera ett framgångsrikt argument som leder till denna slutsats.

Projektet utförs i två steg, vilka överlappar med avhandlingens två delar. Del I, som omfattar kapitel 2–4, utforskar vad exakt det är med våra moraliska uppfattningars historia som ger oss skäl att avfärda dem. Mer specifikt innebär det ett sökande efter vilka principer som skulle kunna förklara hur genealogisk information kan undergräva trovärdigheten hos våra moraliska uppfattningar. Detta görs genom en detaljerad granskning av tre mycket omdiskuterade undergrävande förklaringar och genom att försöka vaska fram allmänna principer ur dem.

I Del II, som omfattar kapitel 5–8, utvärderas de olika principerna som presenterats i Del I, genom att undersöka utsikterna för de undergrävande förklaringar som använder dem. I utvärderingen kommer jag framför allt att fokusera på två saker. För det första kommer jag att undersöka olika strategier för att bemöta debunking argument. För det andra kommer jag att titta närmare på interna utmaningar som sådana argument står inför, varav jag primärt kommer fokusera på dessa fyra: behovet av plausibla epistemiska principer, hotet om generalisering, hotet om självundergrävning och nödvändigheten av metaepistemiska ställningstaganden.

I det första och inledande kapitlet börjar jag med att presentera ämnet för avhandlingen—debunking-argument—och ger en kortfattad översikt över litteraturen i ämnet. Därefter ger jag en översikt över olika typer av debunking-argument samt den nödvändiga teoretiska bakgrunden och de epistemologiska ramar som jag använder i resterande kapitel. Därefter diskuterar jag vilka teorier om moraliskt tänkande och talande som har störst anledning att oroa sig över att få debunking-argument riktat mot sig. I samband med detta motiverar jag också vissa begränsningar av räckvidden för mitt centrala argument. Slutligen ger jag en översikt över de kommande kapitlen och lyfter fram mina bidrag till litteraturen om debunking-argument.

I kapitel 2–4 undersöks hur debunking-argument utnyttjar genealogiska faktorer för att globalt undergräva moraliska uppfattningar. Detta görs genom att granska tre klassiska moraliska debunking-argument av Gilbert Harman (kapitel 2), Richard Joyce (kapitel 3) och Sharon Street (kapitel 4). Jag visar att dessa argument inte lyckas förklara hur deras genealogiska påståenden leder till de föregivna epistemologiska slutsatserna. Jag diskuterar därefter en rad epistemiska principer som skulle göra det möjligt för dessa argument att generera slutsatsen att ingen moralisk uppfattning är berättigad. Bland de principer som identifieras finns sådana som rör ontologisk sparsamhet, förklaringsmässig umbärlighet, epistemisk okänslighet, avsaknad av epistemisk trygghet, oförklarlig tillförlitlighet och epistemiska sammanträffanden.

I kapitel 5–8 utvärderas olika debunking-argument som bygger på de olika epistemiska principerna som identifierades i Del I, eller på förfiningar av dessa principer. I kapitel 5 utforskas den så kallade tillförlitlighetsutmaningen för moraliska uppfattningar, enligt vilken de är föremål för en problematisk form av epistemisk tillfällighet och att deras tillförlitlighet är oförklarlig och därför undergrävs. Jag hävdar att en viktig komponent i tillförlitligheten hos moraliska uppfattningar är förklaringen på det jag kallar den *minimala korrelationen*. Detta är sambandet mellan ens egna moraliska uppfattningar och de fakta som man anser att dessa uppfattningar handlar om. Detta samband behöver förklaras genom att förklara den relevanta korrelationen. Jag presenterar en taxonomi för hur man kan förklara dessa korrelationer, och skiljer mellan fyra olika förklaringsmodeller: (i) direkta förklaringar, som när en faktor kausalt förklarar en annan; (ii) indirekta förklaringar, som när en tredje faktor förklarar båda de korrelerade faktorerna; (iii) slumpmässiga förklaringar, som när två faktorer korrelerar av ren slump; och (iv) fullblods-platonism, som hävdar att våra moraliska uppfattningar garanteras vara korrekta så länga den moraliska teorin vi håller med om är konsistent.

Efter att ha redogjort för tillförlitlighetsutmaningen och möjliga sätt att besvara den på, närstuderar jag försök att ge en indirekt förklaring av den minimala korrelationen genom det som kallas tredjefaktors-förklaringar. Jag hävdar att det finns en förvirring i litteraturen om tredjefaktors-strategin. När strategin korrekt förstås som påståendet att det finns en tredje faktor som spelar en dubbel förklaringsroll—gentemot både moraliska fakta och moraliska övertygelser—så är strategin oanvändbar för många moraliska realister. Utmaningen gällande tillförlitlighet är därför fortfarande ett reellt hot.

I kapitel 6 diskuterar jag ett annat sätt på vilket man kan försöka besvara tillförlitlighetsutmaningen, nämligen genom en förklaring av den minimala korrelationen som innebär att den kan sägas vara en ren slump. Denna strategi—som jag kallar den slumpmässiga korrelationsstrategin—accepterar att sambandet mellan moraliska fakta och våra uppfattningar om dem saknar en enhetlig förklaring. Den accepterar därför att moraliska uppfattningar är tillförlitliga, i den mån de är det, endast av en slump. Trots detta hävdar denna strategi att detta faktum inte är tillräckligt för att undergräva trovärdigheten hos våra moraliska uppfattningar, eftersom även den slumpmässiga förklaringen uppfyller vad som rimligen krävs för att ha tillförlitliga, välgrundade uppfattningar och kunskap.

Jag hävdar att detta svar inte är lika lätt att tillbakavisa som tredjefaktorsstrategin, och att svaret faktiskt verkar kunna lyckas, åtminstone i princip. För närvarande verkar det därför möjligt att blockera de debunking-argument som vi har undersökt genom att acceptera att våra moraliska övertygelser är tillförlitliga, och att det är till följd av en ren slump.

I kapitel 7 diskuterar jag en ny våg av debunking-argument som försvarar villkor för vad som krävs för att ha rationella uppfattningar. Dessa villkor innebär att det måste finnas något förklaringssamband mellan moraliska fakta och moraliska uppfattningar. Om dessa argument lyckas, skulle de kunna blockera strategier som den slumpmässiga korrelationsstrategin. Jag presenterar en ny invändning mot denna typ av debunking-argument och hävdar att de står inför ett dubbelt och sammanflätat hot. För det första hävdar jag att sådana villkor för rationell tro hotar att generalisera på ett sådant sätt att de även kommer utesluta vardagliga och till synes oklanderliga uppfattningar, så som uppfattningar om framtiden. För det andra hotas sådana argument genom att de riskerar att undergräva sig själva. Detta eftersom det inte alls är uppenbart att en tro på villkoret som argumentet vilar på uppfyller villkoret själv. En förespråkare av argumentet måste alltså ta på sig en omfattande förklaringsbörda för att undvika självundergrävning. Särskilt kommer detta innebära att man måste förklara metaepistemiska sakförhållanden, såsom vilka förklaringsmässiga relationer som råder mellan våra epistemiska uppfattningar och epistemiska fakta. Detta är svåra och kontroversiella frågor som riskerar att utplåna all dragningskraft hos det resulterande debunking-argumentet.

Kapitel 8 sammanfattar lärdomarna från de föregående kapitlen och ger en diagnos av varför debunking-argument står inför så svåra hinder. För det första hävdar jag att ett moraliskt debunking-argument måste vila på substantiella epistemiska principer. Detta behövs för att ett sådant argument ska kunna förklara hur våra uppfattningar inom ett visst område undergrävs på ett systematiskt sätt. Sådana kandidater till principer är mycket omtvistade och möter ofta utmaningar på rent epistemologiska grunder. Behovet av sådana principer drar därför en förespråkare av *moraliska* debunking-argument ur den moraliska domänen och in i djupa epistemologiska debatter.

Även om en förespråkare av debunking-argument skulle hitta en trovärdig epistemisk princip som kan fungera som grund för sin argumentation, är arbetet långt ifrån klart. Jag visar att debunking-argument som använder epistemiska principer tenderar att stå inför en allvarlig och underskattad utmaning: sådana argument är antingen självundergrävande eller så måste de försvara kontroversiella metaepistemiska ställningstaganden. För att undvika att vara självundergrävande måste förespråkare av debunking-argument därför ta på sig mycket omtvistade och kontroversiella åsikter inom ett antal områden utanför det område de riktar sig mot. Beroende på vilken typ av argument det rör sig om kan det krävas att de förbinder sig till särskilda epistemologiska uppfattningar inom områden utöver det som de vill avfärda, till exempel områden som matematik, modalitet, logik, metafysik och epistemologi (då alltså metaepistemiska uppfattningar).

En annan överraskande slutsats i avhandlingen är att det finns en betydande skillnad mellan hur två olika typer av debunking-argument kan bemöta utmaningarna jag har tecknat. Den första typen av argument, som syftar till att undergräva våra uppfattningar inom ett visst område helt och hållet, stöter på de svåraste problemen. De argument som bara vill att vi skall avfärda en specifik (icke-skeptisk) metaetisk teori, till fördel för en annan, är den överlägset mest lovande formen av debunking-argument. Det ligger dock i sakens natur att just denna form av argument syftar till att i slutändan ge oss sanna, motiverade uppfattningar om domänerna i fråga. Paradoxalt nog verkar det därför som att just debunking-argument har de ljusaste utsikterna i den grad de egentligen inte syftar till att undergräva våra faktiska uppfattningar.

I korthet hävdar jag att den förklaringsbörda som förespråkarna av debunking-argument står inför har allvarligt underskattats. När man klarlägger den teoretiska helheten som krävs för att framställa ett lyckat debunking-argument, riskerar det att göra debunking-argumentet mycket mindre tilltalande än vad som tidigare antagits. Det krävs ett grundligt och omfattande arbete från förespråkare av debunking-argument om de önskar att framställa argument som håller vad de utlovar.

För att föra litteraturen vidare tecknar jag två adekvansvillkor för alla framtida debunking-argument. Dessa gör det möjligt för sådana argument att navigera hoten från generalisering och självunderminering. Jag skisserar också några framtida riktningar som förespråkare av debunking-argument måste följa för att återupprätta utsikterna för ett framgångsrikt globalt moraliskt debunking-argument.

References

Alexander, David. 2011. "In Defense of Epistemic Circularity." *Acta Analytica* 26 (3): 223–41.

Alfano, Mark. 2016. *Moral Psychology: An Introduction*. Polity.

Allhoff, Fritz. 2003. "Evolutionary Ethics from Darwin to Moore." *History and Philosophy of the Life Sciences* 25 (1): 51–79.

Alston, William P. 1988. "The Deontological Conception of Epistemic Justification." *Philosophical Perspectives* 2: 257–99.

Armstrong, D. M. 1973. *Belief, Truth and Knowledge*. London: Cambridge University Press.

Arvan, Marcus. 2021. "Morality as an Evolutionary Exaptation." In *Empirically Engaged Evolutionary Ethics*, edited by Johan De Smedt and Helen De Cruz, 89–109. Springer - Synthese Library.

Balaguer, Mark. 1998. *Platonism and Anti-Platonism in Mathematics*. Oxford University Press.

Baras, Dan. 2017a. "Our Reliability Is In Principle Explainable." *Episteme*

———. 2017b. "A Reliability Challenge to Theistic Platonism." *Analysis* 77 (3): 479–87.

———. 2020. "A Strike against a Striking Principle." *Philosophical Studies* 177: 1501–14.

Barker, Jonathan. 2020. "Debunking Arguments and Metaphysical Laws." *Philosophical Studies* 177 (7): 1829–55.

Barnes, E. 2000. "Ockham's Razor and the Anti-Superfluity Principle." *Erkenntnis* 53 (3): 353–74.

Baron, Sam. 2017. "Feel the Flow." *Synthese* 194 (2): 609–30.

Becker, Kelly. 2012. "Methods and How to Individuate Them." In *The Sensitivity Principle in Epistemology,* edited by Kelly Becker and Tim Black, 81–98. Cambridge: Cambridge University Press.

Becker, Kelly, and Tim Black. 2012. *The Sensitivity Principle in Epistemology*. Cambridge University Press.

Bedke, Matthew S. 2009. "Intuitive Non-Naturalism Meets Cosmic Coincidence." *Pacific Philosophical Quarterly* 90 (2): 188–209.

———. 2014. "No Coincidence?" In *Oxford Studies in Metaethics,* edited by Russ Shafer-Landau, 9:102–25. Oxford University Press.

Behrends, Jeff. 2013. "Meta-Normative Realism, Evolution, and Our Reasons to Survive: Meta-Normative Realism, Evolution." *Pacific Philosophical Quarterly* 94 (4): 486–502.

Benacerraf, Paul. 1973. “Mathematical Truth.” *Journal of Philosophy* 70 (19): 661–79.

Bergmann, Michael. 2006. *Justification without Awareness: A Defense of Epistemic Externalism*. Oxford : New York: Clarendon Press ; Oxford University Press.

Berker, Selim. 2014. “Does Evolutionary Psychology Show That Normativity Is Mind-Dependent?” In *Moral Psychology and Human Agency*, edited by Justin D’Arms and Daniel Jacobson, 215–52. Oxford University Press.

———. 2019a. “The Explanatory Ambitions of Moral Principles.” *Noûs* 53 (4): 904–36.

———. 2019b. “The Explanatory Ambitions of Moral Principles.” *Noûs* 53 (4): 904–36.

Björklund, Fredrik, Jonathan Haidt, and Scott Murphy. 2000. *Moral Dumbfounding: When Intuition Finds No Reason*. Vol. Vol 1. Lund Psychological Reports 2. Department of Psychology, Lund University.

Bogardus, Tomas. 2016. “Only All Naturalists Should Worry About Only One Evolutionary Debunking Argument.” *Ethics* 126 (3): 636–61.

Bogardus, Tomas, and Chad Marxen. 2014. “Yes, Safety Is in Danger.” *Philosophia* 42 (2): 321–34.

Boghossian, Paul. 2006. *Fear of Knowledge: Against Relativism and Constructivism*. Oxford University Press.

BonJour, Laurence. 1998. *In Defense of Pure Reason*. Cambridge University Press.

———. 2010. “Recent Work on the Internalism-Externalism Controversy.” In *A Companion to Epistemology, Second Edition*, edited by Jonathan Dancy, Ernest Sosa, and Matthias Steup, 33–43. Blackwell.

Boyd, Richard. 1988. “How to Be a Moral Realist.” In *Essays on Moral Realism*, edited by G. Sayre-McCord, 181–228. Cornell University Press.

Braddock, Matthew. 2017. “Debunking Arguments from Insensitivity.” *International Journal for the Study of Skepticism* 7 (2): 91–113.

Brink, David O. 1989. *Moral Realism and the Foundations of Ethics*. Cambridge Studies in Philosophy. Cambridge ; New York: Cambridge University Press.

Brogaard, Berit, and Joe Salerno. 2013. “Remarks on Counterpossibles.” *Synthese* 190 (4): 639–60.

Brosnan, Kevin. 2011. “Do the Evolutionary Origins of Our Moral Beliefs Undermine Moral Knowledge?” *Biology & Philosophy* 26 (1): 51–64.

Carter, J. Adam, and Robin McKenna. 2021. “Absolutism, Relativism and Metaepistemology.” *Erkenntnis* 86: 1139–59.

Case, Spencer. 2018. “From Epistemic to Moral Realism.” *Journal of Moral Philosophy*, 1–22.

Cecchetto, Cinzia, Raffaella Ida Rumiati, and Valentina Parma. 2017. “Relative Contribution of Odour Intensity and Valence to Moral Decisions.” *Perception* 46 (3–4): 447–74.

Chang, Ruth. 2004. “‘All Things Considered.’” *Philosophical Perspectives* 18 (1): 1–22.

Chrisman, Matthew. 2007. “From Epistemic Contextualism to Epistemic Expressivism.” *Philosophical Studies* 135 (2): 225–54.

Christensen, David. 2010. “Higher-Order Evidence.” *Philosophy and Phenomenological Research* 81 (1): 185–215.

Clarke-Doane, Justin. 2015. “Justification and Explanation in Mathematics and Morality.” In *Oxford Studies in Metaethics*, edited by Russ Shafer-Landau, 10:80–103. Oxford University Press.

———. 2016. “Debunking and Dispensability.” In *Explanation in Ethics and Mathematics: Debunking and Dispensability*, edited by Uri D. Leibowitz and Neil Sinclair, 23–36. Oxford University Press.

———. 2017. “Debunking Arguments: Mathematics, Logic, and Modal Security.” In *The Cambridge Handbook of Evolutionary Ethics*, edited by Michael Ruse and Robert J. Richards, 202–9. Cambridge: Cambridge University Press.

———. 2019. “Modal Objectivity.” *Noûs* 53 (2): 266–95.

———. 2020. *Morality and Mathematics*. Oxford University Press.

Comesaña, Juan. 2005. “Unsafe Knowledge.” *Synthese* 146 (3): 395–404.

———. 2010. “Evidentialist Reliabilism.” *Noûs* 44 (4): 571–600.

Copp, David. 1990. “Explanation and Justification in Ethics.” *Ethics* 100 (2): 237–58.

———. 2008. “Darwinian Skepticism about Moral Realism.” *Philosophical Issues* 18 (1): 186–206.

Cosmides, L., R. A. Guzmán, and J. Tooby. 2018. “The Evolution of Moral Cognition.” In *Routledge Handbook of Moral Epistemology*, edited by Aaron Zimmerman, Karen Jones, and Mark Timmons.

Côté-Bouchard, Charles, and Clayton Littlejohn. 2018. “Knowledge, Reasons, and Errors about Error Theory.” In *Metaepistemology: Realism & Antirealism*, edited by Robin McKenna and Christos Kyriacou. Palgrave Macmillan.

Cowie, Christopher. 2019. *Morality and Epistemic Judgement: The Argument From Analogy*. 1st ed. Oxford University Press.

Crouch, Margaret A. 1993. “A ‘Limited’ Defense of the Genetic Fallacy.” *Metaphilosophy* 24 (3): 227–40.

Crow, Daniel. 2016. “Causal Impotence and Evolutionary Influence: Epistemological Challenges for Non-Naturalism.” *Ethical Theory and Moral Practice* 19 (2): 379–95.

Cuneo, Terence. 2003. “Moral Explanations, Minimalism, and Cognitive Command.” *The Southern Journal of Philosophy* 41 (3): 351–65.

———. 2007. *The Normative Web: An Argument for Moral Realism*. Oxford University Press.

Cuneo, Terence, and Christos Kyriacou. 2018. “Defending the Moral/Epistemic Parity.” In *Metaepistemology*, edited by C. McHugh J. Way and D. Whiting.

Cuneo, Terence, and Russ Shafer-Landau. 2014. “The Moral Fixed Points: New Directions for Moral Nonnaturalism.” *Philosophical Studies* 171 (3): 399–443.

Das, Ramon. 2016. “Evolutionary Debunking of Morality: Epistemological or Metaphysical?” *Philosophical Studies* 173 (2): 417–35.

DeRose, Keith. 1995. “Solving the Skeptical Problem.” *Philosophical Review* 104 (1): 1–52.

Devitt, Michael. 2005. “There Is No a Priori.” In *Contemporary Debates in Epistemology*, edited by Steup Matthias and Sosa Ernest, 105–15. Blackwell.

———. 2011. “No Place for the A Priori.” In *What Place for the A Priori?*, edited by Michael J. Shaffer and Michael Veber, 9–32. Open Court.

DiPaolo, Joshua. 2018. “Higher-Order Defeat Is Object-Independent.” *Pacific Philosophical Quarterly* 99 (2): 248–69.

Douven, Igorn D. 2011. “Abduction.” *Stanford Encyclopedia of Philosophy*.

Dretske, Fred. 1971. “Conclusive Reasons.” *Australasian Journal of Philosophy* 49 (1): 1–22.

Dunaway, Billy. 2017. “LUCK: EVOLUTIONARY AND EPISTEMIC.” *Episteme* 14 (4): 441–61.

Dworkin, Ronald. 1996. “Objectivity and Truth: You’d Better Believe It.” *Philosophy & Public Affairs* 25 (2): 87–139.

Dyke, Michelle M. 2020. “Bad Bootstrapping: The Problem with Third-Factor Replies to the Darwinian Dilemma for Moral Realism.” *Philosophical Studies* 177 (8): 2115–28.

Egeland, Jonathan. 2022. “The Epistemology of Debunking Argumentation.” *The Philosophical Quarterly*, January, pqab074.

Eklund, Matti. 2017. *Choosing Normative Concepts*. Oxford University Press.

———. 2020. “The Normative Pluriverse.” *Journal of Ethics and Social Philosophy* 18 (2).

Elga, Adam. 2010. “How to Disagree about How to Disagree.” In *Disagreement*, edited by Ted Warfield and Richard Feldman, 175–86. Oxford University Press.

Enoch, David. 2009. “Can There Be a Global, Interesting, Coherent Constructivism about Practical Reason?” *Philosophical Explorations* 12 (3): 319–39.

———. 2010. “The Epistemological Challenge to Metanormative Realism: How Best to Understand It, and How to Cope with It.” *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 148 (3): 413–38.

———. 2011. *Taking Morality Seriously: A Defense of Robust Realism*. Oxford University Press.

———. 2019. “How Principles Ground.” *Oxford Studies in Metaethics* 14.

Enoch, David, and Tristram McPherson. 2017. “What Do You Mean ‘This Isn’t the Question’?” *Canadian Journal of Philosophy* 47 (6): 820–40.

Enoch, David, and Joshua Schechter. 2008. “How Are Basic Belief-Forming Methods Justified?” *Philosophy and Phenomenological Research* 76 (3): 547–79.

Faraci, David. 2015. “A Hard Look at Moral Perception.” *Philosophical Studies* 172 (8): 2055–72.

———. 2019. “Groundwork for an Explanationist Account of Epistemic Coincidence” 19 (4): 26.

Feldman, Richard. 2005. “Respecting the Evidence.” *Philosophical Perspectives* 19 (1): 95–119.

Feldman, Richard, and Earl Conee. 2001. “Internalism Defended.” *American Philosophical Quarterly* 38 (1): 1–18.

Field, Hartry. 1989. *Realism, Mathematics & Modality*. Blackwell.

———. 2005. “Recent Debates about the A Priori.” In *Oxford Studies in Epistemology Volume 1*, edited by Tamar Szabo Gendler and John Hawthorne. Vol. 1. Oxford University Press.

Finlay, Stephen. 2007. “Four Faces of Moral Realism.” *Philosophy Compass* 2 (6): 820–49.

———. 2014. *Confusion of Tongues: A Theory of Normative Language*. Oxford University Press.

FitzPatrick, William J. 2015. “Debunking Evolutionary Debunking of Ethical Realism.” *Philosophical Studies* 172 (4): 883–904.

———. 2016. “Misidentifying the Evolutionary Debunkers’ Error: Reply to Mogensen.” *Analysis* 76 (4): 433–37.

———. 2020. “Morality and Evolutionary Biology.” In *Stanford Encyclopedia of Philosophy*, edited by Edward Zalta.

Fogal, Daniel, and Olle Risberg. 2020. “The Metaphysics of Moral Explanations.” *Oxford Studies in Metaethics* 15.

Fraser, Benjamin James. 2014. “Evolutionary Debunking Arguments and the Reliability of Moral Cognition.” *Philosophical Studies* 168 (2): 457–73.

Freud, Sigmund. 1927. *The Future of an Illusion*. Broadview Press.

Fumerton, Richard A. 1995. *Metaepistemology and Skepticism*. Studies in Epistemology and Cognitive Theory. Lanham, Md: Rowman & Littlefield.

Garner, Richard, ed. 2019. *The End of Morality: Taking Moral Abolitionism Seriously*. New York: Routledge.

Gaut, Berys. 2007. *Art, Emotion and Ethics*. Oxford University Press.

Gettier, Edmund L. 1963. “Is Justified True Belief Knowledge?” *Analysis* 23 (6): 121–23.

Gödel, Kurt. 1964. “What Is Cantor’s Continuum Problem?” In *Philosophy of Mathematics: Selected Readings*, 52:258–73. Englewood Cliffs, N.J., Prentice-Hall.

Goldman, Alvin I. 1967. “A Causal Theory of Knowing.” *Journal of Philosophy* 64 (12): 357–72.

———. 1976. “Discrimination and Perceptual Knowledge.” *The Journal of Philosophy* 73 (20): 771–91.

———. 1979. “What Is Justified Belief?” In *Justification and Knowledge: New Studies in Epistemology*, edited by George Sotiros Pappas, 1–23. Dordrecht: Springer Netherlands.

———. 2011. “Toward a Synthesis of Reliabilism and Evidentialism? Or: Evidentialism’s Troubles, Reliabilism’s Rescue Package.” In *Evidentialism and Its Discontents*, edited by Trent Dougherty, 254–80. Oxford University Press.

Goldman, Alvin I., and Bob Beddor. 2021. “Reliabilist Epistemology.” In *Stanford Encyclopedia of Philosophy*.

Golub, Camil. 2017. “Expressivism and the Reliability Challenge.” *Ethical Theory and Moral Practice* 20 (4): 797–811.

Gowans, Christopher W. 2000. “Introduction.” In *Moral Disagreements: Classic and Contemporary Readings*, edited by Christopher W Gowans, 1–43. Routledge.

Greco, Daniel. 2015. “Epistemological Open Questions.” *Australasian Journal of Philosophy* 93 (3): 509–23.

Greco, John. 2012. “Better Safe than Sensitive.” In *The Sensitivity Principle in Epistemology*, edited by Kelly Becker and Tim Black, 193–206. Cambridge: Cambridge University Press.

Greene, Joshua D. 2008. “The Secret Joke of Kant’s Soul.” In *Moral Psychology, Vol. 3*, edited by W. Sinnott-Armstrong, 35–79. MIT Press.

———. 2013. *Moral Tribes: Emotion, Reason and the Gap Between Us and Them*. Penguin Press.

———. 2016. “Solving the Trolley Problem.” In *Solving the Trolley Problem*, edited by Justin Sytsma and Wesley Buckwalter, 175–78. Malden, MA: Wiley Blackwell.

Greene, Joshua D., Fiery A. Cushman, Lisa E. Stewart, Kelly Lowenberg, Leigh E. Nystrom, and Jonathan D. Cohen. 2009. “Pushing Moral Buttons: The Interaction between Personal Force and Intention in Moral Judgment.” *Cognition* 111 (3): 364–71.

Grundmann, Thomas. 2009. “Reliabilism and the Problem of Defeaters.” *Grazer Philosophische Studien* 79 (1): 65–76.

———, ed. 2011. “Defeasibility Theory.” In *The Routledge Companion to Epistemology*, 156–66. New York: Routledge.

Haidt, Jonathan. 2001. “The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment.” *Psychological Review* 108 (4): 814–34.

Haig, Brian. 2007. “Spurious Correlation.” In *Encyclopedia of Measurement and Statistics*, edited by Neil J. Salkind, 937–40.

Hamilton, W.D. 1964a. “The Genetical Evolution of Social Behaviour. I.” *Journal of Theoretical Biology* 7 (1): 1–16.

———. 1964b. “The Genetical Evolution of Social Behaviour. II.” *Journal of Theoretical Biology* 7 (1): 17–52.

Handfield, Toby. 2016. “Genealogical Explanations of Chance and Morals.” In *Explanation in Ethics and Mathematics: Debunking and Dispensability,* edited by Uri D. Leibowitz and Neil Sinclair. Oxford University Press.

Harman, Gilbert. 1965. “The Inference to the Best Explanation.” *Philosophical Review* 74 (1): 88–95.

———. 1977. *The Nature of Morality: An Introduction to Ethics*. New York: Oxford Univ. Press.

———. 1984. “Is There a Single True Morality?” In *Morality, Reason and Truth: New Essays on the Foundations of Ethics,* edited by David Copp and David Zimmerman, 27–48. Rowman & Allanheld.

———. 1986a. *Change in View*. MIT Press.

———. 1986b. “Moral Explanations of Natural Facts-Can Moral Claims Be Tested Against Moral Reality?” *The Southern Journal of Philosophy* 24 (1): 57–68.

Harman, Gilbert, and Judith Jarvis Thomson. 1996. *Moral Relativism and Moral Objectivity*. Blackwell.

Heathwood, Chris. 2009. “Moral and Epistemic Open-Question Arguments.” *Philosophical Books* 50 (2): 83–98.

———. 2018. “Epistemic Reductionism and the Moral-Epistemic Disparity.” In *Metaepistemology: Realism & Antirealism,* edited by Christos Kyriacou and Robin McKenna, 45–70. Palgrave Macmillan.

Hetherington, Stephen. 1998. “Actually Knowing.” *The Philosophical Quarterly (1950-)* 48 (193): 453–69.

Hopster, Jeroen. 2019. “Striking Coincidences: How Realists Should Reason about Them.” *Ratio* 32 (4): 260–74.

Huemer, Michael. 2001. *Skepticism and the Veil of Perception*. Lanham: Rowman & Littlefield.

———. 2005. *Ethical Intuitionism*. New York: Palgrave Macmillan.

———. 2007. “Compassionate Phenomenal Conservatism.” *Philosophy and Phenomenological Research* 74 (1): 30–55.

———. 2014. “Phenomenal Conservatism.” In *The Internet Encyclopedia of Philosophy,* edited by James Fieser and Bradley Dowden. http://www.iep.utm.edu/phen-con/.

———. 2016. “A Liberal Realist Answer to Debunking Skeptics: The Empirical Case for Realism.” *Philosophical Studies* 173 (7): 1983–2010.

Ichikawa, Jonathan, and Matthias Steup. 2017. “The Analysis of Knowledge.” In *Stanford Encyclopedia of Philosophy*.

Isserow, Jessica. 2019. “Evolutionary Hypotheses and Moral Skepticism.” *Erkenntnis* 84: 1025–45.

Jackson, Frank. 1998. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford University Press.

Jenkins, C. S. 2008. *Grounding Concepts: An Empirical Basis for Arithmetical Knowledge*. Oxford ; New York: Oxford University Press.

Joyce, Richard. 2001. *The Myth of Morality*. Cambridge, UK ; New York: Cambridge University Press.

———. 2006. *The Evolution of Morality*. Life and Mind. Cambridge, Mass: MIT Press.

———. 2008. "Preçis of 'The Evolution of Morality.'" *Philosophy and Phenomenological Research*, 213–18.

———. 2016a. *Essays in Moral Skepticism*. Oxford University Press.

———. 2016b. "Reply: Confessions of a Modest Debunker." In *Explanation in Ethics and Mathematics: Debunking and Dispensability*, edited by Uri D. Leibowitz and Neil Sinclair, 124–45. Oxford University Press.

Kahane, Guy. 2011. "Evolutionary Debunking Arguments." *Noûs* 45 (1): 103–25.

———. 2014. "Evolution and Impartiality." *Ethics* 124 (2): 327–41.

Kallberg, Luke J. 2021. "Evolution and Knowledge." Saint Louis University.

Kelly, Daniel Ryan. 2011. *Yuck!: The Nature and Moral Significance of Disgust*. Bradford.

Kelly, Thomas. 2008. "Common Sense as Evidence: Against Revisionary Ontology and Skepticism." *Midwest Studies In Philosophy* 32 (1): 53–78.

Kelp, Christoph. 2009. "Knowledge and Safety." *Journal of Philosophical Research* 34: 21–31.

Kim, Jaegwon. 1988. "What Is 'Naturalized Epistemology?'" *Philosophical Perspectives* 2: 381–405.

Kitcher, Philip. 2005. "Biology and Ethics." In *The Oxford Handbook of Ethical Theory*, edited by David Copp, 163–85. Oxford University Press.

———. 2011. *The Ethical Project*. Harvard University Press.

Klement, Kevin C. 2002. "When Is Genetic Reasoning Not Fallacious?" *Argumentation* 16 (4): 383–400.

Klenk, Michael. 2018. "Survival of Defeat: Evolution, Moral Objectivity, and Undercutting." Utrecht University.

———. 2019. "Objectivist Conditions for Defeat and Evolutionary Debunking Arguments." *Ratio* 32 (4): 246–59.

———, ed. 2020. *Higher-Order Evidence and Moral Epistemology*. Routledge Studies in Epistemology. New York: Routledge.

Korman, Daniel Z. 2014. "Debunking Perceptual Beliefs about Ordinary Objects." *Philosophers' Imprint* 14: 1–21.

———. 2019a. "Debunking Arguments." *Philosophy Compass* 14 (12): 1–17.

———. 2019b. "Debunking Arguments in Metaethics and Metaphysics." In *Metaphysics and Cognitive Science*, edited by Alvin I. Goldman and Brian P. McLaughlin, 337–63. New York, NY: Oxford University Press.

Korman, Daniel Z., and Dustin Locke. 2020. “Against Minimalist Responses to Moral Debunking Arguments.” In *Oxford Studies in Metaethics, Volume 15*, 309–32. Oxford University Press.

———. 2021. “An Explanationist Account of Genealogical Defeat.” *Philosophy and Phenomenological Research*, October, phpr.12848.

Kornblith, Hilary. 2008. “Knowledge Needs No Justification.” In *Epistemology: New Essays*, edited by Quentin Smith, 5–23. Oxford University Press.

Kvanvig, Jonathan L. 2003. *The Value of Knowledge and the Pursuit of Understanding*. Cambridge Studies in Philosophy. Cambridge: Cambridge Univ Press.

———. 2007. “Two Approaches to Epistemic Defeat.” In *Alvin Plantinga*, edited by Deane-Peter Baker, 107–24. Cambridge University Press.

Kyriacou, Christos. 2016. “Are Evolutionary Debunking Arguments Self-Debunking?” *Philosophia* 44 (4): 1351–66.

———. 2018. “From Moral Fixed Points to Epistemic Fixed Points.” In *Metaepistemology: Realism & Antirealism*, edited by Christos Kyriacou and Robin McKenna. Palgrave Macmillan.

———. 2019. “Evolutionary Debunking: The Milvian Bridge Destabilized.” *Synthese* 196 (7): 2695–2713.

Lasonen-Aarnio, Maria. 2014. “Higher-Order Evidence and the Limits of Defeat.” *Philosophy and Phenomenological Research* 88 (2): 314–45.

Lazari-Radek, Katarzyna de, and Peter Singer. 2012. “The Objectivity of Ethics and the Unity of Practical Reason.” *Ethics* 123 (1): 9–31.

Lewis, David. 1971. “Immodest Inductive Methods.” *Philosophy of Science* 38 (1): 54–63.

———. 1973. *Counterfactuals*. Blackwell.

Lillehammer, Hallvard. 2003. “Debunking Morality: Evolutionary Naturalism and Moral Error Theory.” *Biology & Philosophy* 18 (4): 567–81.

Locke, Dustin. 2014. “Darwinian Normative Skepticism.” In *Challenges to Moral and Religious Belief: Disagreement and Evolution*, edited by Michael Bergmann and Patrick Kain, 220–36. Oxford University Press.

Lott, Micah. 2018. “Must Realists Be Skeptics? An Aristotelian Reply to a Darwinian Dilemma.” *Philosophical Studies* 175 (1): 71–96.

Lutz, Matt. 2018. “What Makes Evolution a Defeater?” *Erkenntnis* 83 (6): 1105–26.

———. 2020. “The Reliability Challenge in Moral Epistemology.” In *Oxford Studies in Metaethics, Volume 15*, 284–308. Oxford University Press.

Lutz, Matthew, and James Lenman. 2021. “Moral Naturalism.” In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2021. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/spr2021/entries/naturalism-moral/.

Lycan, William G. 2002. “Explanation and Epistemology.” In *The Oxford Handbook of Epistemology*, edited by Paul K. Moser, 413. Oxford University Press.

Mackie, John Leslie. 1977. *Ethics: Inventing Right and Wrong*. Penguin Books.

Maguire, Barry, and Jack Woods. 2020. “The Game of Belief.” *The Philosophical Review* 129 (2): 211–49.

Majors, Brad. 2007. “Moral Explanation.” *Philosophy Compass* 2 (1): 1–15.

Mantel, Susanne. 2019. “Do Epistemic Reasons Bear on the Ought Simpliciter?” *Philosophical Issues* 29 (1): 214–27.

Mayr, Ernst. 1961. “Cause and Effect in Biology.” *Science* 134 (3489): 1501–6.

McCain, Kevin. 2015. “Explanationism: Defended on All Sides.” *Logos and Episteme* 6 (3): 333–49.

McHugh, Conor, Jonathan Way, and Daniel Whiting. 2018a. *Metaepistemology*. Oxford University Press.

———, eds. 2018b. *Normativity: Epistemic and Practical*. First edition. Oxford, United Kingdom: Oxford University Press.

McPherson, Tristram. 2011. “Against Quietist Normative Realism.” *Philosophical Studies* 154 (2): 223–40.

Merricks, Trenton. 2003. “Replies.” *Philosophy and Phenomenological Research* 67 (3): 727–44.

Mogensen, Andreas L. 2014. “Evolutionary Debunking Arguments in Ethics.” Dissertation, Oxford.

———. 2015. “Evolutionary Debunking Arguments and the Proximate/Ultimate Distinction.” *Analysis* 75 (2): 196–203.

Moon, Andrew. 2017. “Debunking Morality: Lessons from the EAAN Literature: Debunking Morality.” *Pacific Philosophical Quarterly* 98: 208–26.

Moretti, Luca, and Tommaso Piazza. 2018. “Defeaters in Current Epistemology: Introduction to the Special Issue.” *Synthese* 195 (7): 2845–54.

Nagel, Thomas. 2012. *Mind and Cosmos: Why the Materialist Neo-Darwinian Conception of Nature Is: Why the Materialist Neo-Darwinian Conception of Nature Is Almost Certainly False*. Oxford University Press USA.

Neta, Ram, and Guy Rohrbaugh. 2004. “Luminosity and the Safety of Knowledge.” *Pacific Philosophical Quarterly* 85 (4): 396–406.

Nichols, Shaun. 2014. “Process Debunking and Ethics.” *Ethics* 124 (4): 727–49.

Nolan, Daniel. 2013. “Impossible Worlds.” *Philosophy Compass* 8 (4): 360–72.

Nozick, Robert. 1981. *Philosophical Explanations*. Harvard University Press.

Oddie, Graham. 2005. *Value, Reality, and Desire*. Clarendon Press.

Olin, Lauren, and John M. Doris. 2014. “Vicious Minds: Virtue Epistemology, Cognition, and Skepticism.” *Philosophical Studies* 168 (3): 665–92.

Olson, Jonas. 2019. “What Can Debunking Do for Us (Sceptics and Nihilists)?” *Ratio,* February.

Parfit, Derek. 2011. *On What Matters, Vol 2*. The Berkeley Tanner Lectures. Oxford ; New York: Oxford University Press.

Plantinga, Alvin. 1993. *Warrant and Proper Function*. New York: Oxford University Press.

———. 2000. *Warranted Christian Belief*. Oxford University Press.

Pollock, John L. 1986. *Contemporary Theories of Knowledge*. Hutchinson.

Price, Huw, and Brad Weslake. 2008. “The Time-Asymmetry of Causation.” In *The Oxford Handbook of Causation*, edited by Helen Beebee, Peter Menzies, and Christopher Hitchcock, 414–43. Oxford University Press.

Pritchard, Duncan. 2005. *Epistemic Luck*. Vol. 29. Oxford University Press UK.

———. 2007. “Anti-Luck Epistemology.” *Synthese,* 277–97.

———. 2009. “Safety-Based Epistemology: Whither Now?” *Journal of Philosophical Research* 34: 33–45.

———. 2012. “Anti-Luck Virtue Epistemology.” *Journal of Philosophy* 109 (3): 247–79.

Pryor, James. 2004. “What’s Wrong with Moore’s Argument?” *Philosophical Issues* 14 (1): 349–78.

Pust, Joel. 2001. “Against Explanationist Skepticism Regarding Philosophical Intuitions.” *Philosophical Studies* 106 (3): 227–58.

Rabinowitz, Dani. n.d. “The Safety Condition for Knowledge.” In *Internet Encyclopedia of Philosophy*. Accessed March 16, 2020. https://www.iep.utm.edu/safety-c/#SH3b.

Railton, Peter. 1986. “Moral Realism.” *Philosophical Review* 95 (2): 163–207.

Rea, Michael C. 2002. *World without Design: The Ontological Consequences of Naturalism*. Oxford : New York: Clarendon Press ; Oxford University Press.

Ridge, Michael. 2007. “Epistemology for Ecumenical Expressivists.” *Aristotelian Society Supplementary Volume* 81 (1): 83–108.

Rini, Regina A. 2016. “Debunking Debunking: A Regress Challenge for Psychological Threats to Moral Judgment.” *Philosophical Studies* 173 (3): 675–97.

Risberg, Olle, and Folke Tersman. 2019. “A New Route from Moral Disagreement to Moral Skepticism.” *Journal of the American Philosophical Association* 5 (2): 189–207.

———. 2020. “Disagreement, Indirect Defeat, and Higher-Order Evidence.” In *Higher Order Evidence and Moral Epistemology,* edited by Michael Klenk, 97–114.

———. ms. “When Higher-Order Evidence Matters, and Why.”

Roojen, Mark van. 2015. *Metaethics: A Contemporary Introduction*. Routledge.

Rosenberg, Alex. 2012. *The Atheist’s Guide to Reality*. W. W. Norton & Company.

Rowland, Richard. 2013. “Moral Error Theory and the Argument from Epistemic Reasons.” *Journal of Ethics and Social Philosophy* 7 (1): 1–24.

———. 2021. *Moral Disagreement*. New Problems of Philosophy. New York City: Routledge.

Royzman, Edward B, Kwanwoo Kim, and Robert F Leeman. 2015. “The Curious Tale of Julie and Mark: Unraveling the Moral Dumbfounding Effect.” *Judgment and Decision Making* 10 (4): 18.

Ruse, Michael. 1986. *Taking Darwin Seriously*. Prometheus Books.

Ruse, Michael, and Edward O. Wilson. 1986. “Moral Philosophy as Applied Science.” *Philosophy* 61 (236): 173–92.

Russell, Bertrand. 1919. *Introduction to Mathematical Philosophy*. Dover Publications.

———. 2009. *Human Knowledge: Its Scope and Limits*. Routledge Classics. London: Routledge.

Sampson, Eric. 2019. “The Self-Undermining Arguments from Disagreement.” In *Oxford Studies in Metaethics*, 14:23–46. Oxford University Press.

Sauer, Hanno. 2018. *Debunking Arguments in Ethics*. Cambridge University Press.

Scanlon, T M. 2014. *Being Realistic About Reasons*. Oxford University Press.

Schafer, Karl. 2010. “Evolution and Normative Scepticism.” *Australasian Journal of Philosophy* 88 (3): 471–88.

Schaffer, Jonathan. 2019. “Cognitive Science and Metaphysics: Partners in Debunking.” In *Cognitive Science and Metaphysics*, edited by Alvin I. Goldman and Brian P. McLaughlin, 38–69. New York, NY: Oxford University Press.

Schechter, Joshua. 2010. “The Reliability Challenge and the Epistemology of Logic.” *Philosophical Perspectives* 24 (1): 437–64.

———. 2013a. “Could Evolution Explain Our Reliability about Logic?” In *Oxford Studies in Epistemology 4*, edited by Tamar Szabo Gendler and John Hawthorne, 4:214.

———. 2013b. “Rational Self-Doubt and the Failure of Closure.” *Philosophical Studies* 163 (2): 429–52.

———. 2018a. “Explanatory Challenges in Metaethics.” In *Routledge Handbook of Metaethics*, edited by Tristram McPherson and David Plunkett, 443–59. Routledge.

———. 2018b. “Is There a Reliability Challenge for Logic?” *Philosophical Issues* 28 (1): 325–47.

Schwitzgebel, Eric, and Fiery Cushman. 2015. “Philosophers’ Biased Judgments Persist despite Training, Expertise and Reflection.” *Cognition* 141: 127–37.

Shafer-Landau, Russ. 2003. *Moral Realism: A Defence*. Oxford : New York: Clarendon Press ; Oxford University Press.

———. 2006. “Ethics as Philosophy : A Defense of Ethical Nonnaturalism.” In *Metaethics After Moore*, edited by Terry Horgan and Mark Timmons. Oxford University Press.

———. 2007. “Moral and Theological Realism: The Explanatory Argument.” *Journal of Moral Philosophy* 4 (3): 311–29.

———. 2012. “Evolutionary Debunking, Moral Realism and Moral Knowledge.” *Journal of Ethics and Social Philosophy* 7 (1): 1–37.

Siegel, Susanna. 2012. “Cognitive Penetrability and Perceptual Justification*: Cognitive Penetrability and Perceptual Justification.” *Noûs* 46 (2): 201–22.

Sinnott-Armstrong, Walter. 2006. *Moral Skepticisms*. Oxford University Press.

Skarsaune, Knut Olav. 2011. “Darwin and Moral Realism: Survival of the Iffiest.” *Philosophical Studies* 152 (2): 229–43.

Sober, Elliott. 1984. *The Nature of Selection: Evolutionary Theory in Philosophical Focus*. Vol. 95. University of Chicago Press.

———. 1994. *From a Biological Point of View: Essays in Evolutionary Philosophy*. Cambridge Studies in Philosophy and Biology. Cambridge: Cambridge University Press.

———. 2001. “Venetian Sea Levels, British Bread Prices, and the Principle of the Common Cause.” *The British Journal for the Philosophy of Science* 52 (2): 331–46.

———. 2015. *Ockham’s Razors: A User’s Manual*. Cambridge University Press.

Sosa, Ernest. 1999. “How to Defeat Opposition to Moore.” *Philosophical Perspectives* 13: 137–49.

———. 2009. *A Virtue Epistemology: Apt Belief and Reflective Knowledge, Volume I*. Vol. 69. Oxford University Press.

Srinivasan, Amia. 2015. “The Archimedean Urge.” *Philosophical Perspectives* 29 (1): 325–62.

Stalnaker, Robert C. 1968. “A Theory of Conditionals.” In *Studies in Logical Theory (American Philosophical Quarterly Monographs 2)*, edited by Nicholas Rescher, 98–112. Blackwell.

Steiner, Mark. 1973. “Platonism and the Causal Theory of Knowledge.” *Journal of Philosophy* 70 (3): 57–66.

Steup, Matthias. 2013. “Does Phenomenal Conservatism Solve Internalism’s Dilemma?” In *Seemings and Justification: New Essays on Dogmatism and Phenomenal Conservatism*, edited by Chris Tucker, 135. Oxford University Press.

Stevens, Martin. 2013. *Sensory Ecology, Behaviour, and Evolution*. Oxford University Press.

Street, Sharon. 2006. “A Darwinian Dilemma for Realist Theories of Value.” *Philosophical Studies* 127 (1): 109–66.

———. 2008a. “Constructivism about Reasons.” In *Oxford Studies in Metaethics, Volume 3,* edited by Russ Shafer-Landau, 207–45. Oxford Studies in Metaethics. Oxford University Press.

———. 2008b. “Reply to Copp: Naturalism, Normativity, and the Varieties of Realism Worth Worrying About.” *Philosophical Issues* 18 (1): 207–28.

———. 2009. “Evolution and the Normativity of Epistemic Reasons.” *Canadian Journal of Philosophy* 39 (1): 213–48.

———. 2011. “Mind-Independence Without the Mystery: Why Quasi-Realists Can’t Have It Both Ways.” In *Oxford Studies in Metaethics, Volume 6,* edited by Russ Shafer-Landau, 6:1–32. Oxford University Press.

———. 2016. “Objectivity and Truth: You’d Better Rethink It.” In *Oxford Studies in Metaethics*, edited by Russ Shafer-Landau, 11:293–334. Oxford University Press.

Sturgeon, Nicholas L. 1985. “Moral Explanations.” In *Morality, Reason and Truth,* edited by David Copp and David Zimmerman, 49–78.

———. 2006. “Moral Explanations Defended.” In *Contemporary Debates in Moral Theory,* edited by James Dreier, 241–62. Blackwell.

Sudduth, Michael. n.d. “Defeaters in Epistemology.” In *Internet Encyclopedia of Philosophy*. Accessed February 25, 2020. https://www.iep.utm.edu/ep-defea/#SH1b.

Swain, Stacey, Joshua Alexander, and Jonathan M. Weinberg. 2008. “The Instability of Philosophical Intuitions: Running Hot and Cold on Truetemp.” *Philosophy and Phenomenological Research* 76 (1): 138–55.

Sytsma, Justin, and Wesley Buckwalter, eds. 2016. *A Companion to Experimental Philosophy*. Malden, MA: Wiley Blackwell.

Talbott, William J. 2015. “How Could a ‘Blind’ Evolutionary Process Have Made Human Moral Beliefs Sensitive to Strongly Universal, Objective Moral Standards?” *Biology & Philosophy* 30 (5): 691–708.

Tersman, Folke. 2017. “Debunking and Disagreement.” *Noûs* 51 (4): 754–74.

Thurow, Joshua C. 2013. “The Defeater Version of Benacerraf’s Problem for a Priori Knowledge.” *Synthese* 190 (9): 1587–1603.

Tiberius, Valerie. 2014. *Moral Psychology: A Contemporary Introduction*. Routledge.

Tolhurst, William. 1987. “The Argument from Moral Disagreement.” *Ethics* 97 (3): 610–21.

Tomasello, Michael. 2016. *A Natural History of Human Morality*. Harvard University Press.

Tropman, Elizabeth. 2014. “Evolutionary Debunking Arguments: Moral Realism, Constructivism, and Explaining Moral Knowledge.” *Philosophical Explorations* 17 (2): 126–40.

Turri, John. 2010. “On the Relationship between Propositional and Doxastic Justification.” *Philosophy and Phenomenological Research* 80 (2): 312–26.

Vavova, Katia. 2014. “Debunking Evolutionary Debunking.” In *Oxford Studies in Metaethics,* 9:76–101. Oxford University Press.

———. 2015. “Evolutionary Debunking of Moral Realism.” *Philosophy Compass* 10 (2): 104–16.

Vogel, Jonathan. 2012. “The Enduring Trouble with Tracking.” In *The Sensitivity Principle in Epistemology*, edited by Kelly Becker and Tim Black, 122–51. Cambridge: Cambridge University Press.

Waal, F. B. M. de. 2006. *Primates and Philosophers: How Morality Evolved*. The University Center for Human Values Series. Princeton, N.J: Princeton University Press.

Wang, Jennifer. 2021. “The Epistemological Objection to Modal Primitivism.” *Synthese* 198 (April): 1887–98.

Warenski, Lisa. 2021. “Epistemic Norms: Truth Conducive Enough.” *Synthese* 198 (3): 2721–41.

Warner, Rebecca M. 2013. *Applied Statistics: From Bivariate through Multivariate Techniques, 2nd Ed.* Thousand Oaks, CA, US: Sage Publications, Inc.

Weigel, Chris. 2011. “Distance, Anger, Freedom: An Account of the Role of Abstraction in Compatibilist and Incompatibilist Intuitions.” *Philosophical Psychology* 24 (6): 803–23.

Weinberg, Jonathan M., Joshua Alexander, Chad Gonnerman, Shane Reuter, and The Hegeler Institute. 2012. “Restrictionism and Reflection: Challenge Deflected, or Simply Redirected?” Edited by Sherwood J. B. Sugden. *Monist* 95 (2): 200–222.

Weinberg, Jonathan M., Shaun Nichols, and Stephen Stich. 2001. “Normativity and Epistemic Intuitions.” *Philosophical Topics,* 29 (1–2): 429–60.

Werner, Preston J. 2018. “Moral Perception without (Prior) Moral Knowledge.” *Journal of Moral Philosophy* 15 (2): 164–81.

———. 2020. “Moral Perception.” *Philosophy Compass* 15 (1).

White, Roger. 2006. “Problems for Dogmatism.” *Philosophical Studies* 131 (3): 525–57.

———. 2010. “You Just Believe That Because….” *Philosophical Perspectives* 24 (1): 573–615.

Wielenberg, Erik J. 2014. *Robust Ethics: The Metaphysics and Epistemology of Godless Normative Realism*. Oxford, United Kingdom: Oxford University Press.

———. 2016. “Ethics and Evolutionary Theory.” *Analysis* 76 (4): 502–15.

Wielenberg, Erik J. 2010. “On the Evolutionary Debunking of Morality.” *Ethics* 120 (3): 441–64.

Wilkins, John S., and Paul E. Griffiths. 2012. “Evolutionary Debunking Arguments in Three Domains: Fact, Value, and Religion.” In *A New Science of Religion*, edited by James Maclaurin Greg Dawes. Routledge.

Williamson, Timothy. 2000. *Knowledge and Its Limits*. Oxford ; New York: Oxford University Press.

———. 2007. *The Philosophy of Philosophy*. The Blackwell/Brown Lectures in Philosophy 2. Malden, MA: Blackwell Pub.

Wilson, Edward O., and Michael Ruse. 1985. “Evolution of Ethics.” *New Scientist*, 1985, 102: 1478 edition.

Woods, Jack. 2018. “The Authority of Formality.” In *Oxford Studies in Metaethics*, edited by Russ Shafer-Landau, 13:207–29. Oxford University Press.

Woolfolk, Robert L. 2013. “Experimental Philosophy: A Methodological Critique.” *Metaphilosophy* 44 (1–2): 79–87.

Wright, Crispin. 1992. *Truth and Objectivity*. Harvard University Press.

Index

Are our moral beliefs shaped by evolution rather than objective truths? If so, it is no coincidence that such things as harm reduction, children, and social cooperation are of prime importance in our moral lives. It is exactly what they would be, if we were creatures belonging to a lineage whose ancestors' moral attitudes had been significantly influenced by evolutionary selection pressures.

Could such an origin story for our moral beliefs, when fully spelled out and given empirical backing, provide us with information that would force us—on pain of irrationality or some other epistemic vice—to give them up? And if so, why should we think that evolutionary influences are the only form of murky origins that can undermine our beliefs? There are surely countless other subterranean and veiled influences that similarly affect what we believe—cultural and social affiliation, historical period, gender, religion, upbringing, and so forth. Does the influence from such factors similarly have the power to undermine our beliefs?

This thesis explores the nature of debunking arguments and their implications for our understanding of morality and epistemology.

Conrad Bakka
Conrad Bakka is a PhD candidate in Practical philosophy at Stockholm University. His main research interests include metaethics, error theory, and the epistemology of debunking arguments.

ISBN 978-91-8014-230-4

Department of Philosophy

Stockholm
University