ABSTRACT

Title of Dissertation:	SOME EPISTEMOLOGICAL AND PRACTICAL CHALLENGES TO MORAL REALISM: EVOLUTIONARY DEBUNKING, OVERGENERALIZATION, AND AFTERWARD
	Jimmy Alfonso Licon, Doctor of Philosophy, 2019
Dissertation directed by:	Professor Peter Carruthers, Department of Philosophy

In this dissertation, I examine epistemological and practical challenges to robust moral realism – the view that moral facts are independent of actual or idealized minds, and causally inert. Following an introductory chapter, in the next two chapters, I examine an epistemic challenge to moral knowledge (on moral realism) emanating from 'evolutionary debunking arguments' (EDAs). In the second chapter, I argue that capacity approaches are more plausible than content approaches in that the (i) capacity approach is a more pernicious threat to moral realism; and, (ii) the content approach faces a greater

explanatory burden. In the third chapter, I argue that the overgeneralization objection to EDAs – they *viciously* overgeneralize to domains like the epistemic – faces a dilemma: either EDAs don't overgeneralize as there is an independent reason to trust our beliefs in such non-moral domains; or, they benignly overgeneralize to non-moral domains, if we lack an independent reason, *and evolution would plausibly be distorting*, in that domain. Either way, EDAs don't *viciously* overgeneralize.

In the last chapter, I evaluate moral fictionalism: the view that we have practical reasons to think and act morally (e.g. it enhances self-control), despite holding skeptical or deflationary metaethical views. I argue that there are good philosophical and empirical reasons to think that (a) discarding beliefs is far harder than fictionalists claim; and, (b) robust moral dispositions one would need to effectively think and act morally would inculcate belief, *pace* moral fictionalism. Finally, I argue that keeping moral beliefs mitigates moral risk: there is a live epistemic possibility that (a) we could be wrong in our skeptical or deflationary metaethical views, and (b) if our views about such matters are mistaken, but we act on them, we risk acting seriously wrongly. This is another practical reason to think and act morally. And we must be motivated to act morally to mitigate moral risk – so we should preserve our moral beliefs. So, we have practical reasons to keep our moral beliefs, instead of morally pretending.

Each chapter is written so that it is independent of all of the others.

SOME EPISTEMOLOGICAL AND PRACTICAL CHALLENGES TO MORAL REALISM: EVOLUTIONARY DEBUNKING, OVERGENERALIZATION, AND AFTERWARD

by

Jimmy Alfonso Licon

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2019

Advisory Committee:
Distinguished University Professor Peter Carruthers, Chair
Associate Professor Tomas Bogardus
Assistant Professor Brian Kogelmann
Associate Professor Dan Moller
Professor Christopher W. Morris
Associate Professor Karol Soltan

Dedication

For Sue, and Josie – the most important women in my life.

Acknowledgements

I am deeply indebted to a great many people for helping me overcome as much as I have – my story is miraculous, and it would have been impossible without the love and support of too many people to name. My greatest debt is to my loved ones who supported me during the many years and struggles it took to get here. My mother, Sue, has been a constant source of love that I would be lost without. She instilled in me a sense of perseverance, and a love of knowledge, which has carried me through many storms in life. My life partner, Josie, has been a recent, and much needed source of support – her ability to learn and grow inspires me to never give up. I look forward to many years at her side, and I can't imagine life without her (nor do I want to).

My other debt is to Peter Carruthers, my academic advisor. His patience, work ethic, and fastidious attention to rigor and detail have made me a better philosopher – and there is no doubt that his example will continue to be a source of inspiration and growth.

Also, I want to thank Louise Gilman, who has been a friend and source of support and solace during my time at the University of Maryland. Her dedication to everyone in the philosophy department is invaluable, and too often overlooked.

Next, I want to thank my friend and mentor, Bernard Molyneux. He is a world class philosopher whom I try to emulate daily in my own philosophical life.

Finally, I want to thank David Lopez who first got me interested in philosophy. Without his initial introduction to philosophy, and encouragement along the way, I doubt that I would have gotten this far.

I've received support, feedback, and encouragement from friends and colleagues: Tomas Bogardus, Brian Kogelmann, Kevin Timpe, Carlos Montemayor, Steven Hales, Dan Moller, Caleb Robichaud, Andrew Fyfe, Evan Westra, Julius Schönherr, Kyley Ewing, and many others (apologies to those whom I've unintentionally missed).

Table of Contents

Dedication
Acknowledgments
Chapter 1: The Epistemology of Normativity – Context and Background……………….1
1. Introduction………………………………………………………………….1
2. What's the Deal With Evolutionary Debunking Arguments?............................3
2.1. Harman's Explanatory Challenge……………………………………5
2.2. Joyce and Explanatory Challenges…………………………………...6
2.3. Challenges to Explanatory Challenges……………………………….8
2.4. Reliability and Moral Realism………………………………………13
2.5. Normative versus Moral Skepticism………………………………...15
3. Does Evolutionary Debunking 'Prove' *Too* Much?..18
3.1. Do EDAs 'Prove' Too Much?..18
3.2. Companions in Guilt Arguments……………………………………18
3.3. The Self-Defeat or Toothless Objection……………………………20
4. Moral Skepticism or Error-Theory … Then What?..25
4.1. The Nature of Belief………………………………………………...26
4.2. The Difficulty in Managing Belief…………………………………26
4.3. Indispensible, False Beliefs…………………………………………29
5. The Bigger Picture and Conclusion…………………………………………...32
Chapter 2: On Evolutionary Debunking – Street versus Joyce…………………………34
1. Introduction……………………………………………………………….....33
2. The Content Approach: Street………………………………………………...34
3. The Capacity Approach: Joyce………………………………………………..37
4. The Limited Explanation Objection…………………………………………..38
4.1. A General Worry……………………………………………………40
4.2. Content vs. Capacity Debunking……………………………………41
4.3. Limited Explanations and Appeals to Truth………………………...43
5. The Independent Confirmation Objection…………………………………….44
5.1. Defeasible and Decisive Evidence…………………………………46
6. The Overgeneralization Objection……………………………………………47
6.1. Overgeneralization and a Dilemma…………………………………48
6.2. Evolutionary Debunking and Perception…………………………...51
7. Conclusion…………………………………………………………………….54
Chapter 3: How Not to Debunk Evolutionary Debunking………………………………55
1. Introduction……………………………………………………………….....55
2. The Overgeneralization Objection……………………………………………57
2.1. Framing the Debate…………………………………………………57
2.2. Three Versions of the Overgeneralization Objection……………….58
3. The Independent Reason Dilemma…………………………………………...60
3.1. The Independent Reason Condition…………………………...…...60
3.2. Overgeneralizing By Analogy……………………………………....63
3.3. Overgeneralizing By Entailment……………………………………...71
3.4. Overgeneralizing By Entanglement………………………………...75

4. Conclusion…………………………………………………………………….79
Chapter 4: Moral Fictionalism and Moral Risk……………………………………….80
1. Introduction…………………………………………………….…………...80
2. Moral Fictionalism………………………………………….…………….82
2.1. Two Alternatives to Moral Fictionalism……………….…………82
2.2. The Purported Benefits of Moral Fictionalism……….………..88
2.3. Problems with Moral Fictionalism…………………….……………90
3. Moral Risk…………………………………………………………….....98
3.1. Where Moral Risk Enters……………………………………………98
3.2. Does Moral Risk Favor an Approach?..107
4. Conclusion………………………………………………………………...108
References……………………………………………………………………….....110

Chapter 1: The Epistemology of Normativity – Context and Background

1 | Introduction

This dissertation is comprised of three chapters about the epistemological and practical implications of moral realism. This view requires some clarification before we proceed. Some philosophers hold that moral realism is the view that (at least) some moral claims are 'literally construed' and 'literally true' (Sayre-McCord 1988: 5). But this version of moral realism is too thin for our purposes in that many debunkers hold that robust moral realism is *particularly* vulnerable to evolutionary debunking and informed disagreement. So, in this dissertation, we focus exclusively on *robust moral realism*: some moral claims (*i*) hold independent of actual or idealized minds; and, (*ii*) are causally inert (hereafter, call this view *moral realism*—Enoch 2011; Wielenberg 2009; Shafer-Landau 2003).

Each paper is concerned with an aspect of moral realism: chapter 2 broadly distinguishes two approaches to evolutionary debunking (the content versus capacity approach), and argues that we should prefer the capacity approach to the alternative. Chapter 3 critically evaluates a serious objection to evolutionary debunking arguments (EDAs) called 'the overgeneralization objection' – I argue that this objection shouldn't worry the debunker. And chapter 4 defends a weighty practical reason to keep our moral beliefs, *pace* moral fictionalism, if one holds a deflationary or skeptical meta-ethical view of moral facts.

The dissertation centers on the issue of what kind of epistemic grip (if any) we have on moral claims if they are realist in nature: why think that given intractable disagreement or our evolutionary history, we have epistemic access to realist moral facts? And if we cannot get an epistemic grip, how should we proceed? These questions are broadly the focus of this dissertation. This chapter is meant to situate the chapters of the dissertation into broader context of evolutionary debunking arguments (EDAs), and the implications of such arguments for other normative pursuits, like epistemology – and then, finally, the practical implications if such arguments are highly plausible.

Here’s how the chapter will proceed. First, we discuss how EDAs relate to explanatory challenges in metaethics. We examine Harman’s explanatory challenge to moral realism, and how it relates to similar EDAs. We flesh out the explanatory challenge a couple ways to understand how it is an epistemological obstacle to moral realism, and then examine evidence that we aren’t reliable, with respect to our moral beliefs and practices, in a way that the moral realist should find worrisome.

Second, we discuss the nature of ‘companions in guilt’ (CIG) arguments. This kind of argument is commonly used in philosophy, especially trying to vindicate the credentials of a contentious domain like, say the moral and modal, by showing how rejecting facts in that domain requires that one reject facts in a less contentious domain like, say, the epistemic and prudential. The idea is that rejecting facts in the former domain is more costly than one might think as it requires rejecting facts in the latter, less contentious, domain. We discuss then whether EDAs are self-defeating or toothless, as a result of their (purported) overreach to non-moral domains.

Finally, we discuss the practical implications of rejecting moral realism. Suppose that we adopt moral skepticism or error-theory – then what? Moral fictionalists hold that we should discard our moral beliefs, and engage in moral pretense; moral conservationists hold that we should preserve them. We then discuss how the empirical evidence that beliefs are easily acquired and hard to discard relates to philosophical issues like Pascal's Wager and the value of useful, but false beliefs.

2 | What's the Deal With Evolutionary Debunking Arguments?

It is tempting to find evolutionary debunking arguments (EDAs) puzzling. For one thing, there is nothing essentially evolutionary about EDAs: we could just as easily appeal to physical facts, cultural practices, and so on, to explain our moral capacity (without appeal to moral facts). And evolutionary explanations aren't *complete* anyway; they may provide an ultimate explanation, but they don't provide a proximate explanation for our moral beliefs. Worse still, evolutionary explanations must still appeal to factors like culture, physical facts, and historical accidents. If so, then why specifically appeal to evolution to undermine the justification of our moral beliefs? One answer is that such explanations offer a 'naturalistic explanation of our moral beliefs, by providing the *ultimate causes* of moral beliefs to supplement the *proximate causes* that are uncovered by sociology and neuropsychology' to rule out the possibility that 'culture and psychology are ultimately explained by mind-independent, non-natural moral truth' (Lutz 2018: 1118—original emphasis).

But why should facts about our evolution be reason to distrust our moral beliefs? This looks like an instance of the genetic fallacy[1]: the mere fact our moral faculties have been shaped by evolutionary processes (like many cognitive faculties) doesn't thereby give us reason to distrust our moral beliefs[2]. As Rachels (1990) explains:

> We cannot, as a general rule, validly derive conclusions about what ought to be the case from premisses about what is the case. Darwin's theory […] concerns matters of fact. It tells us what is the case, with respect to the evolution of species. Therefore, strictly speaking, no conclusion follows from it regarding any matter of value. It does not follow, merely because we are kin to the apes, that we ought to think less of ourselves, that our lives are less important, or that human beings are 'merely' one kind of animal among others (3).

How should the evolutionary debunker reply? She should acknowledge that appeals to evolution, without appropriate metaethical framing, wouldn't do much: the mere fact that we are the product of evolution doesn't undermine the justification of our moral beliefs. For instance, evolutionary debunkers do not endorse *perceptual* skepticism, despite the fact that evolution shaped our perceptual faculties. Likewise evolution, by itself, cannot tell us about the epistemic status of our moral beliefs. In order to see how the evolutionary debunker can bridge the metaethical gap from our evolutionary history to

[1] Jong and Visala argue that EDAs conflate the distinction between 'the context of discovery and the context of justification' (2014).

[2] Here I mean moral beliefs, sentiments, intuitions, and so forth.

metaethical views like moral skepticism and error-theory, we should review the nature of explanatory challenges in metaethics (Schechter 2018).

2.1 | Harman's Explanatory Challenge

Harman defends the explanatory challenge by arguing that 'explaining the observations that support a physical theory, scientists typically appeal to mathematical principles' but 'one never seems to need to appeal in this way to moral principles' (1977: 10). The idea here is that if we can explain all our evidence, without having to posit moral facts, then that is good reason to be skeptical that there are moral facts[3]. For example, if we can explain all witch-beliefs, without positing any witches, then that is reason to eliminate witches from our ontology; positing witches doesn't do any explanatory work. In contrast, physicists have good reason to posit subatomic particles to explain observations like vapor trails in cloud chambers; subatomic particles 'earn their explanatory keep' — moral facts do not. Harman (1977) makes this point as follows:

> Observation plays a role in science that it does not seem to play in ethics. The difference is that you need to make assumptions about certain physical facts to explain the occurrence of the observations that support a scientific theory, but you do not seem to need to make assumptions about any moral facts to explain the occurrence of the so-called moral observations I have been talking about. In the moral case, it would seem that you *need only make assumptions about the psychology or moral sensibility* of the person making the moral observation (6 – emphasis mine).

[3] We should posit facts only if they do some kind of explanatory work.

If we combine the explanatory requirement (i.e. we should posit an entity, *F*, only if we need it to explain observations, *O*), with the claim that we can explain moral observations like, say, the belief that ‘it is wrong to torture babies for fun’ without having to posit moral facts, then that is at least a good reason to be skeptical that there are moral facts – they do not directly or indirectly feature in our total explanation of our moral cognition[4]. Unless we identify moral facts with non-moral facts, then we have explanatory reasons to reject moral facts on grounds that eliminating such facts from our ontology would be more parsimonious, ceteris paribus.

Harman’s challenge is especially troublesome for *robust* moral realism. If, after all, moral facts were causally efficacious, then they could explain our moral cognition by, say, influencing our evolution. Harman holds that we should embrace moral reductionism: moral facts are a subset of non-moral facts – this reduction allows moral facts to do explanatory work. However, if we think that non-moral facts lack features essential to moral facts, this reductionist move would fail. And that is what evolutionary debunkers like Joyce think.

2.2 | Joyce and Explanatory Challenges

Joyce defends an explanatory challenge that appeals to evolution. We can explain our ability to make moral judgments by appealing to evolutionary processes, without positing moral facts. The idea here is that we have a plausible evolutionary account of our moral

[4] Sturgeon (1986) argues that moral facts *do* explanatory work (e.g. that Hitler was morally depraved helps explain Hitler’s action). Harman could reply that it is Hitler’s bigotry and paranoia that explains his actions; such facts do not obviously need to appeal to moral facts. Similar moves apply to Sturgeon’s other examples.

cognition that ‘nowhere implies or presupposes that these beliefs are true’ (Joyce 2008: 217). Let’s review this challenge.

First, Joyce holds that moral thought and action was adaptive: among other things, it helped enhance self-control and facilitate long-term cooperation:

> By providing a framework within which both one’s own actions and others’ actions may be evaluated, moral judgments can act as a kind of “common currency” for collective negotiation and decision making. *Moral judgment thus can function as a kind of social glue, bonding individuals together in a shared justificatory structure and providing a tool for solving many group coordination problems*. Of particular importance is that although a nonmoralized strong negative emotional reaction (e.g., anger) may prompt a punitive response, it takes a moral judgment to license punishment, and thus the latter serves far more effectively to govern public decisions in a large group than do non-moralized emotions, especially when such emotions may (at the end of a long day’s hunting and gathering) be listless, distracted, or divided (Joyce 2006: 117—emphasis mine).

Second, we need not posit moral facts to explain the adaptive value of thinking and acting morally. Moral cognition would have been adaptive, despite whether there are moral facts (Clarke-Doane 2012): we can explain moral beliefs and faculties, with an evolutionary explanation that ‘nowhere implies or presupposes that these beliefs are true’ (Joyce 2008: 217). And, Joyce holds that we can’t reduce moral facts to natural facts as such facts plausibly lack ‘*the inescapable authority* we apparently expect and require of moral values’ where something is a moral fact if it (i) applies, no matter what the ends or desires of a given agent, and (ii) has serious deliberative weight for that all rational agents (Joyce 2006: 191; original emphasis[5]).

But if (i) we do not need to posit moral facts to explain moral cognition, and (ii) moral facts cannot be reduced to naturalistic facts, then we have an explanationist challenge[6] to moral realism: we shouldn't posit moral facts to explain moral cognition in that we can explain them without positing moral facts. The latter move is important for motivating moral skepticism – if naturalistic facts could also be moral facts, then having reason to accept naturalistic facts would be reason to accept moral facts. But this can't be so if Joyce is right that naturalistic facts can't be moral facts.

2.3 | Challenges to Explanatory Challenges

Although we have a handle on Joyce's evolutionary-based explanatory challenge to moral realism, it is unclear how the lack of explanatory work done by positing realist moral facts undermines the justification of our moral beliefs. Why think that our moral beliefs lack justification just because moral facts don't do explanatory work? Saying that moral facts fail to earn their explanatory keep isn't enough. We need bridging principles to move from the fact that moral facts are explanatorily inert to the claim that our moral beliefs are unjustified. Even if the explanatory challenge were true, without such

[5] Joyce (2006: 62) defines 'the inescapable authority of morality' as reasons that (i) apply, regardless of an agent's ends or desires, and (ii) have strong deliberative weight with rational agents (as opposed to etiquette that has the first, but lacks the second).

[6] This is similar to a move made by realists, in response to perceptual skeptics: we should favor our ordinary world beliefs as they better explain our perceptual experiences than skeptical hypotheses (Vogel 2005).

principles, such a move would only rule out explanationist sources of justification for our moral beliefs. And this move alone doesn't rule out coherentist, intuitionist, or reliabilist sources of justification (Sinnott-Armstrong 2006: 44).

Here it is tempting to cash out evolutionary debunking as evidential insensitivity: our moral beliefs aren't evidentially sensitive – had the moral facts been different or absent, our moral beliefs wouldn't reflect that. And if our moral beliefs aren't evidentially sensitive, they aren't justified. Here is where Joyce's approach comes into play:

> On the assumption that my favored hypothesis about the 'moral sense' is correct, it follows that the process by which humans form moral judgements is an unreliable one … Suppose that the actual world contains real categorical requirements … In such a world humans will be disposed to make moral judgments … for natural selection will make it so. Now imagine instead that the actual world contains no such requirements at all … In such a world humans will *still* be disposed to make these judgments … for natural selection will make it so (2001: 162-163).

The argument here relies on evidential sensitivity: we shouldn't trust our moral beliefs because we should expect to have them solely for evolutionary reasons. Epistemologists, however, worry that evidential insensitivity doesn't defeat justification. We should turn to a couple problems with using evidential insensitivity to justify moral skepticism.

A | GENERAL SKEPTICISM

If epistemic justification is formulated as evidential sensitivity, then we face general skepticism: even if deceived by an Evil Demon with false perceptual experiences, we would still hold our ordinary world beliefs. But despite their evidential insensitivity, our ordinary world beliefs are justified. If the evidential insensitivity of our ordinary world beliefs fails to undermine their justification, its absence would also fail to undermine our moral beliefs.

There is another interpretation of Joyce's appeal to evidential insensitivity: we might read Joyce as arguing that moral beliefs are unjustified because they are unreliable. And that this unreliability is a product of their evidential insensitivity: our moral beliefs are unjustified because they are produced by processes that are unreliable. But on a standard conception of reliability, reliability is cashed out in counterfactual terms: the processes that produced moral beliefs are reliable only if they produce mostly true moral beliefs *throughout nearby possible worlds*[7] (and the actual world)[8]. However, framing evolutionary debunking in reliabilist[9] terms leads to a second problem.

B | NECESSARY TRUTHS

Many metaethicists think that basic moral truths hold necessarily – for instance, it is necessarily wrong to torture kids. But if moral truths hold necessarily, then appeals to evidential insensitivity fail: if moral truths hold necessarily, debunkers cannot argue that

[7] Joyce doesn't say what he means by reliability (2001: 162). He plausibly has something like a standard conception of reliability in mind, as in reliability across nearby possible worlds (Goldman 1986).

[8] The fact that our cognitive faculties produce mostly true beliefs *in the actual world* isn't enough to establish that they are reliable – our moral faculties would have to be very lucky in the actual world.

[9] Our moral beliefs are unjustified because they are unreliable, and they are unreliable because they are evidentially insensitive.

our moral beliefs are evidential insensitive – if the moral facts had been different or absent, our moral beliefs would the same for evolutionary reasons. But if basic moral facts hold necessarily, there are no possible worlds where the moral facts are different or absent, but our beliefs are the same.

The issue is that on this conception of reliability, we can't frame evolutionary debunking in terms of unreliability: if moral truths are necessary, there are no nearby possible worlds where we have the same moral beliefs in each world, but where the moral truths are different or absent. We can't justify the unreliability of our moral beliefs by appealing to their evidential insensitivity.

These worries seem to undermine framing EDAs in terms of evidential insensitivity; this suggests that evidential sensitivity is the wrong epistemic tool to use. We turn to a recent approach to evidential insensitivity that isn't susceptible to such worries.

C | A NEW KIND OF EVIDENTIAL SENSITIVITY

Braddock (2017) defends a novel approach to evidential insensitivity that can supplement arguments for moral skepticism, and avoids the problems sketched above. He notes that,

> [The] bulk of our actual moral judgments—especially the more particular moral truths represented by our everyday moral judgments—are contingent upon the empirical facts. For example, the moral truth that I wrongfully lied to someone depends on empirical facts about, say, my deceptive intention and/or the consequences of my lying. Consequentialists, deontologists, and virtue ethicists highlight the dependence of moral truths on various empirical facts. And since empirical

> facts are contingent, the bulk of moral truths represented by our actual moral judgments are contingent (2017: 17).

And,

> [Most] moral truths appear to be fairly contingent rather than modally robust—depending according to normative tastes on, say, consequences, intentions, and what a virtuous person would do … [even if] some general moral judgments' correspond 'to necessary (or modally robust) moral truths, these judgments will be comparatively fewer because the bulk of our moral judgments are more particular (2017: 17).

Does this point revive evidential insensitivity? We read Joyce as arguing that our moral beliefs are unjustified because they are unreliable given their evidentially insensitivity. If we frame moral truths as necessary, then debunking doesn't get off the ground for lack of evidential insensitivity. But given that the bulk of our moral beliefs are contingently true, then this indicates that many moral truths (even if not the *basic* ones) are contingently true too – and evidential insensitivity resurfaces. And then our moral beliefs would be unjustified. Of course, this argument doesn't hold that our moral beliefs *are* produced by unreliable cognitive processes – we cannot rule out this out by pointing to the necessity of basic moral truths.

Finally, how does this new approach to evidential sensitivity avoid general skepticism? The scope of this new version of the insensitivity arguments is *nearby possible worlds*. The basis of the inference to process unreliability, in the revised evidential sensitivity, is enabled by an appeal to insensitivity plus nearby possible worlds (and worlds where skeptical scenarios hold are very distant from the actual world on standard metrics of distance). We turn next to the issue of reliability.

2.4 | Reliability and Moral Realism

Some realists think that our initial moral reasoning has been distorted by evolutionary influences, but we can reason our way to moral knowledge (FitzPatrick 2015; Toner 2011; Enoch 2011). Consider an analogy: there is evidence that our physics intuitions are Aristotelian in nature – presumably because they were good enough for survival (DiSessa 1982). But physicists overcome such biases, and devise physical theories which approximate physical truths. Moral reasoning isn't much different than physics reasoning: if we can correct such biases in the latter case, then we should be able to do the same in the former case. For instance, Enoch thinks that evolutionary processes pushed us toward judgments that cohere with the *pro tanto* goodness of survival-promoting actions – if we couple such judgments with coherentist reasoning, we could, in principle, arrive at moral knowledge, despite distorting evolutionary influences on our moral faculties.

This requires that our moral starting points are reliable enough such that they can serve as inputs to 'coherentist reasoning mechanisms' and 'the resulting moral judgments are produced by a robustly reliable process' (Braddock 2016: 846). However, Braddock holds that there are empirical reasons to hold that our moral starting points are insufficiently reliable to be starting points – resulting instead in a garbage-in-garbage-out problem.

Even if we grant that evolutionary processes likely pushed us toward judgments cohering with the *pro tanto* goodness of survival (FitzPatrick 2015: 888), it is still another matter whether coherentist reasoning could produce moral knowledge. For one thing, nearly every evolvable internally coherent moral system—including non-commonsensical ones—coheres with the weak normative proposition that survival-promoting actions are *pro tanto good, in some cases* (Street 2006: 122; Bedke 2014). We are aware of cross-cultural diversity in pro-social behavior and norms, despite having similar starting point in terms of shared evolutionary history, and similar moral capacities.

There is also good evidence, from high rates of inter- and intra-group violence, that hunter-gatherer societies in the late Pleistocene and early Holocene ages tended to hold nasty cooperative norms that allowed 'morally objectionable collective killing and exploitation of out-group members' in many cases (Braddock 2016: 849). And if we take contemporary moral norms as a benchmark (as moral realists typically do), nasty norms are wildly wrong. If evolutionary processes heavily influenced the content of our moral judgments, then the record of widespread nasty norms is good evidence that nasty norms are easy to evolve. Also, given that in-group and partiality biases dispose us to discount (or oppose) the well-being of outsiders, it is doubtful that an inclusive, expansionist set of moral beliefs would emerge from evolutionary processes. Braddock concludes that the prevalence of nasty norms is strong (but not conclusive) evidence that our moral faculties are not reliable[10]. Coherentist reasoning cannot save us from these garbage-in-garbage-out problems.

2.5 | Normative versus Moral Skepticism

EDAs usually target the justification of our moral beliefs, rather than targeting normative beliefs. But Cline (2018) argues that EDAs targeting moral beliefs are *too* modest, using Joyce's approach to evolutionary debunking to illustrate. He thinks normative skepticism has advantages over moral skepticism.

A | REVISION

Joyce's EDA assumes that revising our moral concepts cannot purge them of error – but if we could, then this might salvage the justification of our (revised) moral beliefs. For instance, we believed that water was an element, not a substance – but we were able to revise our water concept such that we could purge our water beliefs without adopting water skepticism. Similarly, a moral realist might think Joyce's EDA fails to undercut the justification of our moral beliefs, but rather gives us reason to revise our moral beliefs in a way that purges them of error.

An issue in deciding whether to reject or revise the relevant beliefs is the degree to which the revised concept preserves essential features of the original concept. Joyce holds that (i) moral reasons are inescapable, categorical reasons to act; and (ii) there are no such

[10] In the latter part of the paper, Braddock also argues that innate biases (content and context biases) often serve as moral starting points – but they are subject to all kind of irrelevant factors, and greatly influence the moral judgments we make. This is further evidence, in Braddock's estimation, that we aren't reliable about moral matters.

reasons. However, it is open to the revisionary moral realist to instead argue that moral reasons are approximately practical reasons such that we could revise our moral beliefs in way that purges them of error, rather than reject them. For instance, people have lots of reasons to not murder, care for their kids, and so forth. Such reasons arguably preserve lots of what our original moral concepts capture. This point speaks to the plausibility of revising our moral beliefs, instead of purging them.

B | STRONG CONCEPTUAL AND EMPIRICAL CLAIMS

Many philosophers who endorse moral skepticism or error-theory are moved by the claims that (i) moral reasons are authoritative and inescapable, but (ii) no such reasons have these features. Such views hold that moral reasons are authoritative and inescapable; but, if there could be moral reasons without such features, many arguments for moral skepticism and error-theory would fail – if moral reasons lacked objectionable features, then why be skeptical of them?

There is evidence that ordinary moral thought doesn't presuppose that moral reasons are framed as authoritative and inescapable (Beebe and Sackris 2016). While normative thinking is innate – in that individuals can acquire and comply with norms – only some cultures have reasons that they take to be authoritative and inescapable (Machery and Mallon 2010). This evidence, though not decisive, should trouble those who claim that moral reasons are *essentially* inescapable – if they are, then the implication would be that many cultures lack a conception of moral reasons.

C | OVERGENERALIZING

Evolutionary debunkers often focus their efforts either on a subset of moral beliefs (Singer 2005), or moral beliefs across the board (Joyce 2006; Street 2006). One of the problems here is that reasons in favor of such views tend to overgeneralize to claims that the debunker accepts. Joyce argues that we should be skeptical that there are moral reasons, in part, because we would take ourselves to have such reasons – doing so would have adaptive value, whether moral facts exist or not. But if such reasons had been different or absent 'we would still take ourselves to have normative reasons, due to evolutionary forces' (Cline 2018: 155). However, it is inconsistent, to accept this line of thinking, but accept normative reasons (but reject moral ones).

While it might be inconsistent to reject normative skepticism, based on such arguments, adopting normative skepticism might also be implausible. For one thing, it might be self-defeating (Rowland 2013; Husi 2013). After all, if we should be skeptical that there are normative reasons to believe anything, then we should also be skeptical of the reasons in favor of normative skepticism. And that would be self-defeating. We turn to that worry in the next section.

3 | Does Evolutionary Debunking 'Prove' *Too* Much?

Suppose that EDAs are good reason to reject moral realism: evolutionary reasons strongly imply that we needn't posit categorical moral facts to explain our moral cognition. We might still worry that EDAs, although plausible in the moral domain, are too broad: they debunk beliefs in others domains that we think have good justificatory standing.

3.1 | Do EDAs 'Prove' Too Much?

Some critics charge EDAs with overgeneralizing to other epistemically secure domains like, say, the perceptual domain. Rowland (2013) argues that we cannot reject moral beliefs in that they impugn categorical reasons, but still retain epistemic reasons as they are plausibly categorical too:

1. According to the moral error theory, there are no categorical normative reasons.
2. If there are no categorical normative reasons, then there are no epistemic reasons for beliefs.
3. But there are epistemic reasons for belief.
4. So, there are categorical normative reasons (from 2, 3).
5. So, the error theory is false (from 1, 4).

This brings us to the problem of 'companions in guilt.'

3.2 | Companions in Guilt Arguments

So-called 'companions in guilt' (CIG) arguments attempt to show that challenges to the justification of our moral beliefs also (wrongly) undermines the justification of beliefs in non-moral domains like, say, our perceptual beliefs. CIG arguments point out that the supposedly problematic features of a class of claims, A-claims, is shared by a less epistemically contentious class of claims, B-claims, *ceteris paribus* – the fact that A-claims and B-claims share contentious features casts doubt on both kinds of claims, all else being equal. CIG arguments establish this claim either by *entailment*: the premises 'if moral error theory is true, then there are no categorical normative reasons' and 'epistemic reasons are normative reasons' entails that 'if moral error theory, there are no epistemic reasons' Rowland (2013: 1); or *analogy*: if we would have moral beliefs, even if the moral facts had been different or absent for evolutionary reasons, then by analogy, we would also believe we had epistemic reasons, even in the absence of such reasons as 'we would still take ourselves to have [epistemic] reasons, due to evolutionary forces' Cline (2018: 155). If we shouldn't reject B-claims with certain features, then we shouldn't reject A-claims, despite having contentious features in common – having to reject B-claims raises the cost of rejecting A-claims (Lillehammer 2013). The strategy is to show that EDAs are implausible or costly: if they impugn non-moral beliefs in good epistemic standing, they wrongly overreach and should be rejected (even in the moral domain).

Also, we should note that CIG arguments have a *ceteris paribus* clause: if we reject A-claims because of their contentious features like, say, their categorical nature, then we should reject B-claims that share that feature, in the absence of relevant differences between them. If there are such differences, then the fact that A-claims and B-claims share a contentious feature might not be enough to convict them both if, say, A-claims are categorical *and susceptible to distorting evolutionary pressures*, but B-claims are only categorical. (We revisit this issue in section 3.4).

Consider a toy epistemic CIG argument: if we would believe (for evolutionary reasons) that moral claims imply categorical reasons, even if there are no moral reasons, then the same would also apply to what we believe about epistemic reasons: for evolutionary reasons we would believe that there are categorical epistemic reasons, even if they had been different or absent. And epistemic reasons are (at least) prima facie categorical in that ‘the fact that there are dinosaur bones around is a reason for everyone to believe that dinosaurs once roamed the earth, regardless of whether they want to believe this or not’ (Rowland 2013: 3). And the rejection of epistemic reasons *tout court* would hail the end of knowledge. But then EDAs have gone wrong: they should be rejected – they overreach in a costly or implausible way. We next examine whether EDAs are self-defeating or toothless.

3.3 | The Self-Defeat or Toothless Objection

Some critics hold that CIG arguments, given an EDA, give us reason to reject epistemic realism — and they argue that rejecting epistemic realism would be implausible, as it would render EDAs self-defeating or toothless. In this event, we would either have an epistemic reason to believe epistemic nihilism – which would be self-defeating; or, we would lack a reason to believe anything – which would be toothless (Cuneo 2007: 117-8; Kyriacou 2016). If EDAs give us reason to reject epistemic reasons *simplicter*, they are self-defeating or toothless[11]: this is good reason to reject EDAs in that this overreach is costly and implausible. And it raises the cost of adopting moral skepticism.

[11] See Streumer (2013) and Husi (2013).

For instance, Streumer (2013) argues that if error-theory is true, then we don’t have any reason to believe; but this is a virtue of the theory—in part because it helps the error-theorist respond to a number of objections.

There are a couple things to say here.

First, the self-defeat objection only works if epistemic reasons are categorical. However, if epistemic reasons are constructed – only agents with a certain mental constitution have epistemic reasons – then we could reject *categorical* epistemic reasons without rejecting epistemic reasons *altogether*.

The epistemic realist might object that epistemic constructivism doesn't comport with our epistemic intuitions. There are plenty of examples, to which the epistemic realists appeal, where the mere fact that *F* is an epistemic reason to believe that *F*. The truth of epistemic realism explains why the fact that *F* is reason to believe that *F*. And, the epistemic realist continues, we should expect that facts would give epistemic agents of all sorts like, say, extraterrestrials and robots, epistemic reason to believe, whatever their differing desires, mental constitution, and so on.

But there is a constructivist-friendly explanation of such intuitions. Consider Street's point that the epistemic constructivist[12] might regard,

> [Epistemic] reasons as ultimately deriving from that of practical reasons. A constructivist who took this line might argue that, when it comes to the pursuit of many or most of one's ends, one has *instrumental* reasons to have attitudes which one holds accountable to the truth … it's going to be hard to pursue one's end effectively if all one has are pretences and imaginings (245—original emphasis).

[12] For a reply to a prominent objection to epistemic constructivism, see Flowerree (forthcoming).

The idea here is that agents who are capable of belief would also have a practical interest in forming true beliefs. Having true beliefs would help agents achieve their ends. And this is plausibly a feature of all agents — and on this constructivist-friendly explanation, the fact that F would give all agents reason to believe that *F*, without indicating whether the reason was *epistemic* in nature. There are a couple different ways to read the claims that the fact that *F* is a reason to believe that *F*:

a. The fact that *F* is an *epistemic* reason to believe that *F*.
b. The fact that F is a *practical* reason to believe that *F* (given that it would help most, if not all, agents achieve their aims).

Of course, this is not to argue that we *must* accept (b). The point is that (b) can plausibly explain the intuitions and cases that epistemic realists use to motivate their view. And (b) also explains why we intuitively eschew false beliefs. Since (b) is a constructivist-friendly explanation of the intuitions that purportedly motivate epistemic realism, the constructivist can rightly reply that our epistemic intuitions don't obviously undercut her view.

Critics might worry though that epistemic constructivism is implausible or unstable such that the view somehow collapses into epistemic nihilism. But even granting this worry, should we reject EDAs because they imply epistemic nihilism? Critics argue that if EDAs ‘entail an error theory about *epistemic* judgment […] the metaethical arguments have *overreached*. So we should reject them’ (Cowie 2018). This point assumes that,

> If EDAs have costly or implausible consequences – for instance, they overgeneralize to the epistemic domain and self-defeat – then we should reject the EDAs.

Why accept this? The idea here is that if EDAs have costly or implausible consequences, then *there must something wrong with EDAs*. The costly and implausible consequences convict EDAs – as they give us reason to reject them – even if we cannot specify why they fail in the moral domain. There must be something *wrong with EDAs themselves*, if they overreach in a costly or implausible way. But there is an assumption at work that is questionable:

> If EDAs self-defeat in the epistemic domain, then we aren’t justified in applying them to the moral domain either. Call this *the holistic claim*.

Why accept the holistic claim? The fact that EDAs self-defeat in the epistemic domain is reason to think there is something wrong *with the EDA itself*. However, there is another explanation: there are some epistemically relevant differences, between the moral and epistemic domains, which ensure that EDAs will fail in the epistemic domain – without indicating that they fail *simpliciter* like, say, in the moral domain. Consider a helpful analogy: the fact that a heuristic works in a given situation, but fails when used in a different situation, doesn’t convict the heuristic in the former situation – it rather convicts applying it to the new situation. EDAs might work similarly: there are features (perhaps structural) in the epistemic domain which cause trouble for EDAs, but in a manner that doesn't tell us anything about whether they work in the moral domain.

Critics might object that CIG arguments work by entailment or analogy – the alternative explanation ignores the entailment or analogy component of CIG arguments. However, as we said earlier, CIG arguments work only if their *ceteris paribus* condition is satisfied. If the epistemic and moral domains have epistemically relevant difference[13], then the ceteris paribus clause might not be satisfied. But if the epistemic CIG argument succeeds against the evolutionary debunker, she must show that self-defeat in the epistemic domain, for instance, is failure on the part of EDAs, *not a failure with importing them into a domain where they don't belong.* And if so, EDAs might not overreach at all.

4 | Moral Skepticism or Error Theory … Then What?

Suppose that we adopt error-theory or moral skepticism – then what? The last chapter of the dissertation focuses on the practical value of moral thought and action, even given deflationary or skeptical views. The issue of what we should do with our moral beliefs should be framed 'in broadly Humean terms: as some kind of function of individuals' desires' (Joyce forthcoming).

[13] Perhaps the difference is that arguments for *epistemic* nihilism are self-defeating, but arguments for *moral* nihilism aren't (putting aside issues of overgeneralization).

One need not cease thinking and acting in moral terms, despite skeptical or deflationary views. Although it is plausible not to think and act in way discordant with one's ontic view, in some cases there are good reasons to continue thinking and acting in such a way anyway. For example, many recovering addicts cite belief in a higher power and spiritual practices to explain their success, despite their rejecting religious worldviews – and there is empirical evidence for this claim (Heinz et. al., 2010). There are a couple of related questions here: the first is whether there is practical value to thinking and act morally; and given that there is, the second question is whether we should continue thinking and acting in moral terms using beliefs or pretense.

This is where moral fictionalists (Joyce 2001) and moral conservationists (Olson 2017) disagree. Moral fictionalists hold that there are *practical* reasons to think and act morally like improved self-control and enabling long-term cooperation (Joyce 2001: 186; 2006: 111) – the moral conservationist agrees that continued moral practice has practical value. However, they disagree on *how* to secure the practical value of moral thinking: the moral conservationist thinks that we should keep our moral beliefs — as they are useful, even if systemically false – while the moral fictionalist holds that we can get the same practical goodies by morally pretending. Moral fictionalists hold that if we adopt error-theory or moral skepticism, we should discard our moral beliefs – moral conservationists disagree. Part of this dispute is over the nature of belief; we turn to that below.

4.1 | The Nature of Belief

Much of the debate between the moral fictionalist and moral conservationist turns on the nature of belief, and how easily one can form and discard beliefs. For instance, if our moral beliefs are easy to discard, and we can be robustly disposed to think and act morally without moral beliefs, then moral fictionalism wouldn't face doxastic and motivational obstacles. But, if it is hard to discard beliefs, or if beliefs are required to properly motivate moral thinking and action, moral conservationism is more plausible. And this is bad for the moral fictionalist: evidence from cognitive science suggests that it is easier to acquire beliefs, and harder to discard them, than moral fictionalists require.

The model of beliefs used in such studies holds – independent of the debate between moral fictionalists and moral conservationists – that mental states plausibly count as belief if they are inferentially promiscuous: a core feature of beliefs is not only that they involve accepting a proposition, but that they also function as premises in inferences. Beliefs are plausibly mental states that guide our inferential reasoning, and motivate action in concert with mental states like desires and emotions – such mental states look an awful lot like belief.

4.2 | The Difficulty in Managing Belief

There is good evidence that beliefs are easily formed, but hard to discard: experimental evidence shows that subjects under a cognitive load (e.g. subjects who were told to count backwards from 100 in increments of 7) form beliefs merely by entertaining propositions. For instance, something like deliberately not attending to something in one's visual field can induce a cognitive load (Mandelbaum and Quilty-Dunn 2015: 47). The worry here is that self-regulating what we attend to is ubiquitous in everyday life, and this can easily induce a cognitive load. But this is exactly where the moral fictionalist thinks that our moral pretense should operate – where we are most susceptible to forming beliefs.

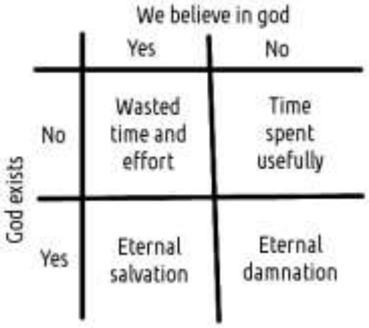
Here's the rub for moral fictionalists: if (i) we have a robust moral sense like, say, having the strong intuition that some actions are morally wrong; and, (ii) simply not attending to our moral intuitions and such is enough to induce a cognitive load (and also make it hard to discard moral beliefs), then by simply consciously not attending to our robust moral sense we would plausibly have a hard time discarding our moral beliefs.

Does this evidence about our doxastic nature inform other debates that turn on the nature of belief? Let's consider an example.

PASCAL'S WAGER

Even if there isn't any evidence that God exists – for the sake of argument, assume that God is equally likely to exist as not – there might still be *practical* reasons to believe in God[14]. This is where prudential arguments in natural theology are salient; Pascal's Wager (hereafter 'the Wager') is a good example. The Wager holds that given the lack of evidence for or against God's existence, we should frame belief in God in terms of expected utility. God promises an infinitely good afterlife to those who believe, and damnation to those who don't:

[14] There are arguably good practical reasons to believe in God *in this life* (McBrayer 2014).



	We believe in god	
God exists	Yes	No
No	Wasted time and effort	Time spent usefully
Yes	Eternal salvation	Eternal damnation

The idea here is that if we believe in God, and we’re right, then the rewards are infinitely good (and goodies in this life too); whereas, if we don’t believe in God, and we’re wrong, then the punishment is infinitely bad. Pascal thinks that expected utility, in the absence of decisive evidence either way, is good reason to believe. Even if we think church is boring (perhaps one would rather watch a football game on Sunday morning), we should still believe in God, as the payoffs (if God exists) would be worth it.

A prominent objection to Pascal’s Wager focuses on the requirement that we *believe* that God exists. And we lack direct control over our beliefs; it’s not as if we can just *decide* to believe in God (Alston 1988). But if lack direct control over our beliefs, then not only does God’s condemnation of our lack of belief seem unfair, but we also wouldn’t reap the reward of believing (correctly) anyway – thus the Wager is a non-starter.

The cognitive science evidence from earlier suggests this objection is wrong (as Pascal’s himself thought): if we put ourselves in the right situation, it wouldn’t be hard to induce something like belief in God. But this move has its own problems: contradictory religious beliefs – including belief that God *doesn't* exist – are also easy to acquire. For instance, if Josie regularly attends church, to induce theistic belief, she must avoid contradictory religious propositions – she should be careful not to read or interact with people critical of her religious beliefs, as this would cause her to believe things contrary to her theistic belief. To the extent that God cares, not only that we believe in Him, but also what *else* we believe about such matters, then the Wager has doxastic troubles – not because it is hard to form theistic beliefs, but because it is all too easy.

4.3 | Indispensible, False Beliefs

Moral fictionalists motivate their view by appealing to the intuition that false beliefs have disvalue. The main reason they offer is that false beliefs would be cognitively onerous as even a 'seemingly useful false belief […] will require all manner of compensating false beliefs to make it fit with what else one knows' (Joyce 2001: 178)[15]. The idea is that if we have a false belief, we must have other false beliefs to make the original false beliefs fit with what we already know. Here's an example from Plantinga (1993):

> Perhaps Paul very much likes the idea of being eaten, but when he sees a tiger, always runs off looking for a better prospect, because he *thinks* it unlikely the tiger he sees will eat him. This will get his body parts in the right place so far as survival is concerned, without involving much by way of true belief ... Or perhaps he *thinks* the tiger is a large, friendly, cuddly pussycat and wants to pet it; but he also *believes* that the best way to pet it is to run away from it (225-6; emphasis mine).

[15] This dovetails nicely with Street's point, in defense of epistemic constructivism, that 'one has *instrumental* reasons to have attitudes which one holds accountable to the truth' (2009: 245).

Here’s an example where a false belief, without compensating beliefs, could result in something very bad for Paul; whereas, if Paul had true tiger-beliefs (and the right desires), then his situation wouldn’t be as cognitively onerous. Paul can avoid the *practical* disvalue of his false-tiger beliefs – but only in a cognitively onerous way. There is lots of practical value to having true beliefs in that they are a better guide to the world than false beliefs, ceteris paribus.

Let’s review an example where false belief has practical value.

IDEALIZATIONS IN SCIENCE

As creatures who engage in science, with limited cognitive resources, and a complex world, we must use idealizations – a model that downplays some, and emphasizes other, features of a phenomenon. Idealizations involve assumptions made without concern for truth; and they are often known to be false. But the complexity of the world we inhabit, and our limited cognitive resources, constrain how we do science. If scientists didn’t use idealizations, their ability to isolate relevant causal factors in the epistemic noise would be lessened -- false idealizations allow them to set aside ‘complicating factors to discern a causal pattern of interest’ (Potochnik 2017: 43). Also the use of idealizations in science is ‘rampant and unchecked’ (Potochnik 2017: 41). Our limited cognitive resources force us to use idealizations that isolate salient features of the world. Idealizations are an essential tool for future science, given the cognitively limited nature of scientists themselves.

A critic might object that idealizations in science need not result in belief; scientists could merely use idealizations to gain insight into a complex issue without believing them. If so, idealizations would *not* be examples of the ubiquity (in such practices) of false, useful beliefs. Rather, this is evidence of the indispensability of false models.

This objection, however, ignores the evidence from cognitive science briefly covered earlier that shows beliefs are easily acquired. Assuming idealizations exhibit 'inferential promiscuity' and motivate us to act on them, then they look a lot like beliefs. And worse still, the cognitive science evidence suggests that 'even when propositions are known to be false, they are passively accepted as soon as they're encoded, and they inferentially integrate with other beliefs' (Mandelbaum and Quilty-Dunn 2014: 45). The claim isn't that we must accept the model of belief underlying such studies; it is that a sufficiently plausible conception of belief, when coupled with other plausible assumptions, is good evidence that false beliefs can be indispensible.

Getting back to the relevance of this discussion for the last chapter of the dissertation: the claim that there are many useful false beliefs undercuts one of the main reasons to favor moral fictionalism over moral conservationism. Moral fictionalists argue that we should rely on moral pretense because moral conservationism requires us to hold systemically false beliefs, and this has practical disvalue. But we've seen that there are cases where, from a practical point of view, we *should* hold false beliefs. There is independent plausibility to the claim, by moral conservationists, that holding false beliefs can have lots of practical value.

5 | The Bigger Picture and Conclusion

In this chapter, we've discussed the epistemic and practical implications of moral realism. First, we reviewed arguments that pose an epistemological challenge to moral realism, called evolutionary debunking arguments (EDAs). We've seen that there are a number of ways to motivate EDAs – and that there are worries having to do with their scope and implications (especially outside the moral domain). We also saw how overgeneralizing worries could be reason to reject EDAs. And why the claim that they overreach in an implausible or costly way might not be worrying. Next, we examined the practical implications of rejecting moral realism, either for skeptical or deflationary reasons, and argued that there are good (but perhaps not decisive) reasons to keep our moral beliefs. And we examined good reasons to think that false beliefs could be very practically value.

There are many avenues in these debates for more research. There is an apparent tension between the claim that our epistemic intuitions can be explained by appealing to the practical value of true beliefs, on the one hand, and the claim that science is good at discovering scientific truths despite the fact that scientists (arguably) use false beliefs to arrive at such. This avenue of research fits nicely with current debates over pragmatic encroachment on reasons for belief (Reisner 2018). And whether the fact that epistemic companions in guilt (CIG) arguments are self-defeating is good reason to think carefully about the conditions under which EDAs are applicable. Such directions for more study are salient to the plausibility of moral realism – and what to do if moral realism is false.

Chapter 2: On Evolutionary Debunking – Street versus Joyce

1 | Introduction

The origins of a belief can undermine its justification like, say, if a belief were produced by an unreliable mechanism. The recent spate of evolutionary debunking arguments (EDAs) operates in this vein: the evolutionary history of our moral capacity threatens to undermine the epistemic standing of our moral beliefs (Street 2006; Joyce 2006; Morton 2016; Fraser 2014). Evolutionary debunking can be formulated as evidential insensitivity (Braddock 2017), or an undercutting defeater (Lutz 2018), among others.

Despite attempts to clarify various approaches to evolutionary debunking, and their metaethical implications (Wielenberg 2016; Vavova 2015), more clarification is needed. To that end, I argue that a capacity approach (Joyce 2006) is better than a content approach (Street 2006), after making the distinction between these approaches. I argue that the capacity approach (*i*) is a more pernicious form of evolutionary debunking, and (*ii*) from an evolutionary point of view, the capacity approach has a lighter explanatory burden than the content approach.

The outline of this paper is this: first, I contrast the content and capacity approaches to evolutionary debunking – on the content approach, selection pressures targeted our basic evaluative attitudes like, say, feeling the pull to help one's kin when they're in distress, which indirectly influences the content of our moral beliefs. Whereas, on the capacity approach, selection pressures targeted our cognitive capacity to make moral beliefs – a cognitive module that issues moral beliefs with embedded moral concepts like, say, desert and ought. Second, I argue that the capacity approach is placed to respond to some standard objections than the content approach; and the capacity approach is more pernicious, and evolutionarily plausible, than the content approach (see section 4.2).

I take *moral realism* to be the view that (a) there are causally inert moral truths that hold independently of actual or idealized minds; and (b) we have moral knowledge of some of them (Shafer-Landau 2009; Parfit 2006). EDAs target the justification of moral beliefs by giving support to conditionals like: *if moral facts are realist in nature, then we lack moral knowledge.* EDAs threaten to undermine moral knowledge if moral facts are realist in nature – and presumably moral realists should find this troubling.

2 | The Content Approach: Street

Street challenges moral realism in the form a dilemma. She begins with the *prima facie* plausible claim that our moral beliefs have been influenced by evolutionary processes:

> My claim is simply that one enormous factor in shaping the *content* of human values has been the forces of natural selection, such that our system of evaluative judgments is thoroughly saturated with evolutionary influence (2006: 114—emphasis mine).

If evolutionary processes heavily influenced our moral beliefs, then either:

1. There *isn't* a relationship between the selection pressures that *indirectly* shaped the content of our moral beliefs and realist moral truths; or,

2. There *is* a relationship between the selection pressures that *indirectly* shaped the content of our moral beliefs and realist moral truths.

Let's consider each in turn.

THE FIRST HORN

Street worries that if our moral beliefs are directed at *mind-independent* moral truths, then they are unlikely to be true in that it is 'possible that as a matter of sheer chance, some large portion of our moral beliefs ended up true' but this 'would require a fluke of luck that's not only extremely unlikely, *in view of the huge universe of logically possible evaluative judgments*' and it would be 'astoundingly convenient to the realist' (122; my emphasis – see also Bedke 2014). It is *possible* a 'large portion' of our moral beliefs are true, but such a situation is highly unlikely given the number of possible of moral beliefs in the normative universe (most of which are false)[16]. We could have had different moral beliefs about, say, the well-being of our kids, our survival, and so on – we have the moral beliefs that we would expect if our moral cognition was heavily influenced by evolution. Street concludes that the moral realist can't opt for the first horn.

[16] Some moral realists (Wielenberg 2009) hold that moral truths are *necessarily* true—but then the claim that there are countless logically possible moral truths would beg the question against the moral realist. Street should cash out her claim in terms of *epistemic* possibilities.

THE SECOND HORN

The moral realist must hold something like ‘the tracking account’ (Street 2006: 125-6), in which we have certain moral beliefs as they were adaptive *and true*, to explain how we could have true moral beliefs despite distorting evolutionary influences. Street thinks, however, that the tracking account has a superior explanatory competitor in ‘the adaptive link account’:

> [Making] certain kinds of evaluative judgements rather than others contributed to our ancestors’ reproductive success not because they constituted perceptions of independent evaluative truths, but rather because they forged adaptive links between our ancestors’ circumstances and their responses to those circumstances, getting them to act, feel, and believe in ways that turned out to be reproductively advantageous (2006: 127).

Street argues that the tracking and adaptive link accounts are explanatory competitors, but that the adaptive link account is preferable as ‘it is more parsimonious; it is much clearer; and it sheds much more light on the explanandum in question’ as ‘human beings tend to make some evaluative judgements rather than others’ (2006: 129). She argues that the adaptive link account better satisfies the relevant explanatory criteria than the tracking account: we need not posit the *truth* of our moral beliefs to explain why we have them.

INDIRECT EVOLUTIONARY INFLUENCES

Street also holds that evolutionary influences on the content of our moral beliefs were *indirect:* evolution favored individuals with specific basic evaluative tendencies that shape shape the content of our moral beliefs. Street argues that ‘rational reflection’ as an independent way to evaluate the truth of our moral beliefs is implausible – rational reflection cannot stand apart from our moral beliefs as ‘rational reflection must always proceed from some evaluative standpoint; it must work from some evaluative premises; it must treat some evaluative judgements as fixed, if only for the time being’ while assessing other evaluative judgements’ (124). Indirect evolutionary influences on our moral cognition shape our moral beliefs. And as we can’t rule this would, we should be skeptical of our moral beliefs.

3 | The Capacity Approach: Joyce

The capacity approach, defended by Joyce, is the view that we have a cognitive module or capacity that embeds moral concepts into our moral beliefs. This module or capacity allows us to think in moral terms because of its evolutionary payoff. The idea is that we have a ‘specialized innate mechanism (or series of mechanisms)’ which ‘comes prepared to categorize the world in morally normative terms; moral *concepts* may be innate, even if moral beliefs are not’ (180-81).

Joyce argues that there is evidence for his view similar to evidence for universal grammar (Dwyer et al. 2010). Moral concepts needn’t be explicitly taught to children; rather, they appear to develop these notions in an extremely reliable sequence. And if children didn’t come equipped with moral concepts like, say, fairness, it is unclear how we could *instill* them (Joyce 2006: 134-36). And the difficult instilling such concepts in children is *prima facie* reason to think that children have an innate moral sense – for a debunking challenge that doesn’t rely on this possibility, see Cline (2015).

Joyce holds that thinking in terms of categorical moral requirements is adaptive:

> By providing a framework within which both one's own actions and others' actions may be evaluated, moral judgments can act as a kind of "common currency" for collective negotiation and decision making. *Moral judgment thus can function as a kind of social glue, bonding individuals together in a shared justificatory structure and providing a tool for solving many group coordination problems*. Of particular importance is that although a nonmoralized strong negative emotional reaction (e.g., anger) may prompt a punitive response, it takes a moral judgment to license punishment, and thus the latter serves far more effectively to govern public decisions in a large group than do non-moralized emotions, especially when such emotions may (at the end of a long day's hunting and gathering) be listless, distracted, or divided (2006: 117—emphasis mine).

In the epistemic component of his debunking argument, Joyce argues that discovering that we have a robust moral module would undermine the justification of our moral beliefs, much like discovering that we had the belief that 'Napoleon was over eight feet tall' merely as a result of taking a belief pill would undermine our Napoleon belief – we can explain the belief without positing an eight foot tall Napoleon. Similarly, if we found that we think in moral terms because we have a moral cognitive module or capacity (regardless of whether there are moral facts), our moral beliefs are epistemically suspect – we have no reason to think they're true (Joyce 2006: 183).

4 | The Limited Explanation Objection

Some argue that EDAs rely on limited explanations of our evaluative thinking (Shafer-Landau 2012; FitzPatrick 2014; Toner 2011): moral cognition, though highly influenced by evolution, can be corrected downstream with tools like, say, rational reflection. And if so, EDAs aren't a potent challenge to the justification of our moral beliefs:

> [Moral judgments] involve employing evaluative and normative concepts in connection with standards and ends, though now conceived as standards and ends defining *what it is to live well all things considered*, rather than just narrow standards of edibility or safety. This is undeniably a major cultural development, requiring abstract thought and training to make the transition to moral judgments that might reliably track moral truths. But that's equally the case for the transition from counting hyenas or estimating the path of a thrown rock to doing research on quantum electrodynamics or modal metaphysics. All of these pursuits require conceptual and analytic sophistication and the use of methodologies that take us far beyond the mental exercises that figured into the evolutionary shaping of our capacities (FitzPatrick 2015: 888-89; original emphasis).

If cognitive tools like abstract reasoning can correct our thinking in other domains – correcting our Aristotelian intuitions in the physics lab, for instance – then, the moral realist argues we can use similar tools to 'build up' reliable, full-fledged moral beliefs that *aren't epistemically suspect*. We can (and often do) correct the distorting influence of evolution on our normative thinking in other areas – even when that influence is pernicious. Unless the debunker can explain why cognitive tools like, say, abstract thinking and rigorous training cannot fix evolutionary distortions on our moral beliefs, just like we can with our scientific beliefs, then the moral realist is right to be confident that at least some of her moral beliefs are justified.

However, critics like Street worry that rational reflection ‘must always proceed from some evaluative standpoint; it must work from some evaluative premises; it must treat some evaluative judgements as fixed, if only for the time being, as the assessment of other evaluative judgements is undertaken’ (2006: 124). If all moral beliefs have been distorted by selection pressures, this process faces the garbage-in-garbage-out problem: it is no solution to compare contaminated moral beliefs with other moral beliefs that are also contaminated. And even if such tools could correct evolutionary distortions in our moral beliefs using correct moral beliefs, there remains the problem of identifying true moral beliefs, and separating them from the distorted moral beliefs.

4.1 | A General Worry

There is a general worry with the limited explanation objection: while there are domains where we can use general reasoning to correct for evolutionary distortions of our starting beliefs (whether moral or not), such domains are either subject to empirical checks like, say, using general reasoning in the physics labs, to correct our Aristotelian intuitions, or they aren’t like, say, with respect to doing metaphysics. FitzPatrick thinks that there is, in principle, no difference between using our general reasoning abilities to fix evolutionary distortions in the moral domain, on the one hand, and distortions in transitions from, say, ‘counting hyenas or estimating the path of a thrown rock to doing research on quantum electrodynamics or modal metaphysics’ (2015: 889).

But practices such as estimating the path of a thrown rock, or doing physics research, can be empirically verified – such practices run up against the world, so to speak. We can empirically evaluate whether the result of such reasoning approximates the truth. But this isn’t obviously so in the moral domain: our ability to do things like evaluate ‘narrow standards of edibility or safety’ doesn’t mean, by itself, that we can use our general reasoning abilities to figure out what it is to live well *all-things-considered* that requires positing realist moral facts.

4.2 | Content vs. Capacity Debunking

The content debunker has a point that evolution likely distorted many of our moral beliefs, but it is unclear how many. Street holds that evolutionary processes distort our basic evaluative tendencies, which in turn indirectly influence our full-fledged moral beliefs. But this leaves open the issue of whether we have basic evaluative tendencies that *aren’t* distorted by evolution – the mere fact that some of our basic evaluative tendencies were distorted by evolution isn’t enough to show that all such evaluative tendencies were distorted (and whether we have moral beliefs based on *those* basic evaluative tendencies). If not, then the moral realist could, in principle, still argue that *those* moral beliefs could count as moral knowledge – a daunting, but still possible, project.

Here the capacity debunker has an advantage over the content debunker in their response to the limited explanation objection. There are a couple of reasons for this.

A | The Embedding Advantage

On the capacity approach, every moral belief is epistemically dubious in that they all run afoul of referential failure – for example, the claim that ‘we ought to keep our promises’ is suspect as the concept of oughtness is embedded in the claim. If every moral belief essentially embeds normative properties we have good evolutionary reasons to distrust, then that is good reason to doubt such judgments. By analogy, if we should be skeptical of the existence of witches, then we *ipso facto* also should be skeptical of any claim that presupposes the existence of witches.

The embedding of suspect moral concept, from a suspect cognitive module, in every moral belief, makes the capacity approach more pernicious than the content approach: if every moral belief has epistemically suspicious moral concepts embedded essentially, every such belief faces is epistemically suspect – evolutionary distortions apply to the entirety of our moral beliefs. On the content approach, in contrast, we could in principle have moral beliefs not based on the basic evaluative tendencies distorted by evolution; this would at least open up the *possibility* that we have moral knowledge, even if it is hard to come by. The capacity approach is a more insidious epistemic challenge to moral realism, *ceteris paribus*: it explains how evolutionary processes distorted every moral belief, given that (*i*) they were produced by a cognitive module with suspect epistemic credentials, and (*ii*) they are embedded with epistemically suspect concepts.

B | The Explanatory Advantage

Here's a worry for the capacity approach: in order for a moral capacity to evolve, we might presume that the majority of moral beliefs that it produces would have to confer adaptive value; otherwise, it isn't clear why *that kind* of moral capacity, rather than one that produced different moral beliefs, evolved. But that is the wrong way to frame the matter. It need only be that there are a few moral beliefs, produced by our moral capacity, which confers adaptive value, so long as it was enough adaptive value overall to justify selecting individuals with that moral capacity, even if most moral beliefs are adaptively neutral, or even maladaptive.

On the content approach, in contrast, the debunker must show that each individual basic evaluative tendency had adaptive value, rather than being maladaptive or adaptively neutral. This is because, on the content approach, basic evaluative tendencies can come apart: we can have some basic evaluative tendencies without others; the content debunker needs an evolutionary story to explain the adaptive value of each and every basic evaluative tendency – or an account of why, if basic tendencies cluster, they cluster that way, rather than clustering differently.

4.3 | Limited Explanations and Appeals to Truth

The limited explanation objection, recall, holds that we can use our general reasoning capacities, and tools like 'abstract reasoning and methodologies' to correct evolutionary distortions of our moral beliefs – the end result could be moral beliefs that don't reflect evolutionary bias (like we do, say, in the physics). But it should be noted that the stories that we tell about using our evolved capacities to do science or mathematics *involves an appeal to the respective truths* to explain such practices (Vavova 2014: 81-2; FitzPatrick 2015). It is plausible that the adaptive value conferred by such abilities requires appealing to mathematical and scientific facts to explain – it matters whether the beliefs in such a domain are true. The capacity to form moral beliefs, however, would *prima facie* have been adaptive, whether there had been any moral facts or not.

But there is another problem: the idea that ‘conceptual and analytic sophistication and the use of methodologies’ (FitzPatrick 2015: 889) allows us to arrive at moral knowledge, despite our distorted moral starting points, can be parodied. Suppose that witch-realists concede to witch-skeptics that there is a plausible genealogical account of our witch-beliefs that neither implies nor presupposes the existence of witches. However, they claim that we can arrive at knowledge of witchcraft using tools like ‘conceptual and analytic sophistication and the use of methodologies.’ But this is an inadequate reply to witch-skeptics: witch judgments are epistemically dubious in that we need not posit actual witches to explain such judgments. Similarly, if moral concepts are epistemically dubious, then every moral belief is likewise epistemically dubious.

5 | The Independent Confirmation Objection

Some realists defend a strategy to vindicate the justificatory status of our moral beliefs indirectly: if they can show that the same faculties that produce moral beliefs are either identical to, or a subset of, faculties that we have independent reason to trust, then we can vindicate our moral beliefs against debunkers indirectly. If faculties that produce non-moral synthetic *a priori* knowledge also generate our moral beliefs, then in virtue of the reliability of the mechanism in non-moral matters, we could have an independent to trust our moral beliefs (Shafer-Landau 2012: 35; Bogardus 2016: 658-9).

Shafer-Landau (2012) expresses the objection as:

> Enter the formal strategy, which seeks to vindicate the reliability of a doxastic faculty by showing that it is identical to, or a species of, a (kind of) doxastic faculty in which we independently have a high degree of warranted confidence. When it comes to our moral faculties, the likeliest candidate to serve as the basis of a successful formal strategy is the faculty of generating a priori beliefs. It may be that there is one such faculty, or an interrelated set of them, that ranges over a wide variety of belief contents. Perhaps the mental operations that generate nonmoral synthetic a priori knowledge are the very ones responsible for generating our a priori moral beliefs. Were that so, we would have good, *albeit defeasible*, reason to consider our moral faculties reliable, at least with regard to the a priori moral beliefs they generate (35—my emphasis).

Suppose that we have non-moral synthetic *a priori* knowledge (hereafter merely *synthetic knowledge*): there are instances of such knowledge like the claims that, say, 'the moral globally supervenes on the non-moral' or 'nothing can be blue all over and green all over at the same time.' Such examples are defeasible reason to trust the faculty responsible. If moral beliefs were produced by a module we had prior reason to trust, then our moral beliefs would be defeasibly justified. And we would have defeasible evidence that our moral beliefs were reliably produced, on this objection.

The debunker here might reply that we cannot 'get outside' our moral beliefs to test their epistemic status. For instance, comparing contaminated moral beliefs against one another won't vindicate our moral beliefs – it just introduces a garbage-in-garbage-out problem (Street 2006: 124). Another issue is that, given evolutionary debunking, we are not *prima facie* entitled to hold some moral beliefs for the purposes of calibration; the distorting influence of evolution is prima facie too pervasive on our moral beliefs for this approach to be plausible. This is the insulation problem.

The independent confirmation objection avoids the insulation problem by taking an indirect route to establish the reliability of our moral faculties. And it also raises the cost of evolutionary debunking: unless the debunker forgoes synthetic knowledge altogether, then this is a *prima facie* independent reason to trust our moral beliefs[17]. If moral beliefs are produced by a cognitive faculty that is reliable in non-moral domains, then we have defeasible evidence that our moral beliefs were produced by a reliable faculty. The moral realist is then able to point to the favorable epistemic track record of the relevant faculties in non-moral domains as independent reason to think that the evolutionary processes that influenced our moral cognition was not wholly off-track with respect to moral beliefs.

5.1 | Defeasible and Decisive Evidence

Even if the faculty that produces moral beliefs is reliable with regard to non-moral beliefs, however, we can resist the independent confirmation objection. It is epistemically possible that in the space of options available to evolution, there was a choice between a synthetic faculty that only and reliably issued synthetic beliefs, non-moral and moral alike (the solitary option); and, one which was only reliable with regard to non-moral synthetic beliefs, but unreliable with regard to moral beliefs (the combination option). A serious possibility here is that selection pressures favor the combination option in that it would confer adaptive value on highly cooperative species, even in the absence of realist

[17] This approach has some similarities with the ‘companions in guilty’ strategy (Cowie 2018).

moral reasons – while also preserving the adaptive value of true, synthetic non-moral beliefs. Creatures who judged that ‘if something is red all over, it cannot be green all over simultaneously’ would likely have enjoyed an adaptive advantage over their counterparts without such a capacity – evolution probably cares about the reliability of our non-moral synthetic beliefs, but not the reliability of our moral beliefs.

There is nothing we know that rules out that selection pressures favored the combination option. Unless the moral realist can show that either (a) there is no such possible faculty in selection space, or (b) such a faculty would have been *less* adaptive than the alternatives, the debunker can accept that we should epistemically trust non-moral synthetic judgments, but that we have evolutionary reasons to expect that such a faculty would be unreliable on moral matters. This is a debunking-friendly explanation as to why we should trust our synthetic faculty only when it produces non-moral beliefs, but not moral beliefs. We can accept that moral beliefs are defeasibly justified in that they are produced by a faulty we have independent reason to trust – but that justification could still be defeated by the evolutionary origins of the synthetic faculty.

6 | The Overgeneralization Objection

Most evolutionary debunking targets the moral domain. However, we might worry that evolutionary debunking overreaches to non-moral domains in a costly or implausible way. And if so, then evolutionary debunking should be rejected. The objection holds that EDAs wrongly overreach to the epistemic domain, for instance, with bad implications, and so shouldn’t be trusted in the moral domain either (Cuneo 2007; Cowie 2018). First, I argue that EDAs either overreach to the epistemic domain benignly, or not at all. Then I argue that EDAs don't overreach to the perceptual domain.

6.1 | Overgeneralization and a Dilemma

Some epistemologists hold that there are realist epistemic reasons to hold certain beliefs like, say, the fact that there are dinosaur bones at the dig is realist epistemic reason to believe that 'there are dinosaur bones at the dig', no matter what we desire (Cuneo 2007; Rowland 2013). The overgeneralization objection worries that EDAs wrongly overreach to the epistemic domain: if concepts like 'ought' are epistemically suspect for evolutionary reasons, then they would also be suspect in the epistemic domain – and that is the conceptual stuff that realist epistemic reasons are made of. And if epistemic realism is more plausible than moral realism – as some moral realists claim – then such overreach raises the cost of evolutionary debunking in the moral domain too; the cost of using evolutionary debunking to defend moral skepticism also includes epistemic skepticism. The idea is to show that evolutionary debunking is too unstable to merely target the moral domain, but leave other domains like, say the epistemic domain, untouched.

But this objection faces the independent reason dilemma: either we have an independent reason to believe that there are categorical epistemic reasons, or we don't.

Debunkers hold that there is no independent reason to trust our moral beliefs – namely, a reason that is outside the domain in question, and free of evolutionary distortion – and critics need an independent reason if they are to avoid begging the question against the debunker[18] (call this *the independent reason condition*). If evolutionary processes would *distort* the beliefs in a given domain, then the independent reason condition becomes especially salient, to be justified in holding beliefs in that domain. As Vavova explains:

> We can now see what the debunker thinks we need if we are to avoid her challenge: a reason to think that we are not mistaken in our evaluative beliefs that doesn't simply presuppose the truth of those beliefs. *This reason is, in some sense, independent of what is called into question.* This explains why the debunker asks us to bracket our evaluative beliefs […] and to focus only on the origin story (2014: 81—emphasis mine).

An independent reason to trust the reliability of our moral faculties would deflate EDAs. Likewise, if we had an independent reason, to trust the beliefs in a given domain, EDAs wouldn't overreach to that domain. However, if we lack an independent reason to trust beliefs in a given domain, where evolution would be distorting, then overreaching to that domain would be benign – or no more troubling than debunking in the moral domain.

A | WE HAVE INDEPENDENT REASON

If there is an independent reason, *in the epistemic domain*, to accept that our beliefs about the nature of *epistemic* oughts are untainted by evolutionary influence, then debunking wouldn't overgeneralize to the epistemic domain – it wouldn't satisfy the no-independent reason requirement of EDAs. Suppose the epistemic realist has an independent reason to think that there are categorical oughts in the *epistemic* domain like, say, if rejecting

[18] Put aside third-factor objections to evolutionary debunking – we cannot explore how the third-factor objection interacts with the independent reason condition (Enoch 2010). For one thing, it is a difficult to determine whether third-factor responses beg the question against the evolutionary debunker (Joyce 2016), or they act like 'defeater-deflectors' (Moon 2017).

epistemic realism would be self-defeating, then there would be an independent reason in the epistemic case, *but not the moral case*, to think that our epistemic beliefs that entail categorical oughts aren't defeated by evolution. So, the evolutionary debunker needn't worry about the charge of overgeneralizing *as it only targeted the moral domain given the conditions specified by the evolutionary debunker*.

It doesn't matter what that independent reason *is*, but rather that there is an independent reason to think that the epistemic domain contains categorical oughts (if we want to prevent overgeneralizing), but not in the moral domain. If there is an independent reason to think we can reliably access realist epistemic facts, then evolutionary debunking *can't* overgeneralize; it fails to overgeneralize to that domain on its own terms.

B | WE LACK AN INDEPENDENT REASON

Suppose that evolution would have been distorting in the epistemic domain[19], but we lack an independent reason to think that there are realist epistemic reason. Then given such conditions, EDAs would seemingly *benignly* overgeneralize – they overgeneralize in a way that is no more worrying than debunking in the moral domain. If the epistemic realist doesn't have an independent reason to think that there are realist oughts in the epistemic domain, then the debunker shouldn't worry: if evolution would have been distorting in the epistemic domain, then without an independent reason to trust such beliefs in that domain, there would be no better reason to think that there are categorical oughts in the

[19] If evolutionary processes would be *enhancing* in a given domain, evolutionary *debunking* wouldn't be applicable to that domain (e.g. if we had to posit facts in given domain to explain the relevant judgments).

epistemic domain than in the moral domain. And the debunker already accepts debunking in the moral domain.

If EDAs are undermining defeaters in the moral domain (Lutz 2018), and we lack an independent reason to posit realist oughts in the epistemic domain, the debunker needn't worry that her approach would defeat her epistemic beliefs that entail realist epistemic oughts any more than she should worry about that her approach debunks beliefs in the moral domain. The moral realist might push back here by arguing that this result would be the end of epistemology – if we should doubt that there are epistemic reasons, then we can't know anything. And this would be too costly.

But that can't be right. For one thing, we could hold that there *are* epistemic oughts, but they are constructed, not categorical, in that, say, Sally ought to believe that *p* given the evidence for *p and her mental make-up* (Street 2009). And while this view of the nature of epistemic oughts may not satisfy the epistemic realist, it is an approach that appears to make room for epistemology, while allowing that there are evolutionary reasons to reject epistemic *realism*. And this isn't the end of the matter – but it is reason to think that we could carry on with epistemology, even if we reject epistemic realism.

6.2 | Evolutionary Debunking and Perception

We might also worry that if we applied a similar debunking challenge to the perceptual domain, then debunking would, by analogy, imply perceptual skepticism. And clearly this implication would be *too* costly[20] – any argument that implies perceptual skepticism must be mistaken. Shaffer-Landau (2012) captures this worry nicely[21]:

> As with morality, evolutionary pressures in the perceptual domain have as their confirmed doxastic effect the cultivation of adaptive dispositions. […] But this effect is not a distorting one, since adaptive perceptual beliefs will, at least for the most part, be true […] I believe that this tempting line of reasoning spells trouble for the debunker. We can know that adaptive perceptual practices are also reliable ones only if we already have a sense of which perceptual judgments are true and which are false. We can tell that dispositions to hold false perceptual beliefs are likely to be *maladaptive only if we can identify some false perceptual beliefs, show that they tend to undermine fitness and make inferences from those cases* […] If we are required to suspend judgment about all perceptual beliefs – *as we must, if required to do so in the moral case* – then we will most likely not be in a position to confirm the reliability of our perceptual faculties. We must presuppose the truth of at least some central, widely uncontroversial perceptual beliefs in order to get the confirmation of our perceptual faculties off the ground […] *we should be given similar license for morality* (21—emphasis mine).

The claim is that the perceptual domain is just as susceptible to evolutionary debunking as the moral domain. And if we accept evolutionary debunking in the moral domain, it would apply to beliefs in the perceptual domain too. The reason is we would need to know that false perceptual experiences would likely be maladaptive – but to do this, we would need to identify false perceptual beliefs, and show them to be maladaptive. The

[20] The fact that perceptual skepticism would undermine the empirical evidence underlining EDAs might makes such arguments self-defeating. If so, this is candidate for an independent reason (in the perceptual domain).

[21] See also Clarke-Doane 2012; Plantinga 1993: Ch. 12

worry is that we can’t simply take certain perceptual beliefs to be true, as way to evaluate the epistemic standing of *other* perceptual beliefs without allowing the moral realist to do likewise. However, if we cannot identify specific false perceptual beliefs, and show they are maladaptive, we cannot argue that evolution would favor reliable perceptual faculties.

The evolutionary debunker might reply that there is an important difference here: having false moral beliefs could have been adaptive, while having false perceptual beliefs would have prima facie been maladaptive. Recall Shaffer-Landau thinks that the evolutionary debunker must identify false perceptual beliefs would generally be maladaptive to appeal to evolutionary considerations to claim that our perceptual faculties would be generally reliable. But, it is unclear why we should accept this constraint. Debunkers don’t argue they can *identify* false, but adaptive moral beliefs; they argue that there is a plausible story she can tell where moral beliefs would be false, but adaptive (Joyce 2006: 24-31), like, say, it is possible that our belief that ‘we ought to ensure the survival of our kids’ is false, but has adaptive value. And we can also tell a plausible story as to why we should expect that evolution would favor individuals with broadly reliable perceptual faculties – we need not identify false maladaptive perceptual beliefs in order to argue that unreliable perceptual faculties would be maladaptive overall.

A few critics aside (Plantinga 1993; Stich 1993), unreliable perceptual faculties wouldn’t be adaptive (Boudry and Vlerick 2014; Deem 2018): creatures must locate food, avoid predators, and so forth, using their perceptual faculties – the simplest way to do this is with reliable perception. The explanatory role played by reliable perception derails the claim that EDAs overreach to perception. And perceptual realists have plausible, though not decisive, replies to external world skeptics that give us reason to trust our perceptual faculties (Huemer 2016; Vogel 1990). Moral realists, however, lack a similar defense; and worse still, there is some empirical evidence that our moral faculties aren't reliable (Braddock 2016). And false moral beliefs could have been adaptive.

7 | Conclusion

I have argued that there are two broad approaches to evolutionary debunking: the content approach exemplified by Street, and the *capacity* approach exemplified by Joyce. The content approach holds that evolution favored individuals with specific basic evaluative tendencies; the capacity approach holds that evolution favored individuals with innate moral mechanisms that categorize the world in moral terms – moral concepts may be innate; but moral beliefs are not. My claim has been that the capacity approach is better able to respond to standard objections in the literature. And the capacity approach is a more pernicious form of debunking that is simpler in evolutionary terms than the content approach (for more, see section 4.2).

Such advantages suggest that debunkers should take a capacity approach to evolutionary debunking for their debunking needs. It is separate issue, of course, whether EDAs are a viable epistemological challenge to moral realism; either way then, the capacity approach is prima facie better than the content approach to debunking, ceteris paribus.

Chapter 3: How Not to Debunk Evolutionary Debunking

1 | Introduction

Evolutionary debunking arguments (EDAs) try to defeat the justification of our moral beliefs by, say, pointing out that we have a plausible explanation of our moral beliefs that 'nowhere implies or presupposes that these beliefs are true' (Joyce 2008: 217) is thought good reason to worry that our evolutionary origins defeat the justification of our moral beliefs – our moral beliefs might lack justification if, say, we would have had the same moral beliefs that we do, even if the moral facts[22] had been different or absent (Braddock 2017). The evolutionary debunker thinks this should worry the moral realist – those who hold that moral facts are independent of minds, actual and idealized, and causally inert (Enoch 2011; Shafer-Landau 2003).

It might help clarify if we formulate an EDA as follows:

> 1. If (a) moral facts are mind-independent, (b) evolution has strongly influenced our moral faculties in a way that defeats the justification of our moral beliefs, and (c) there is no independent confirmation of the reliability of our moral faculties, then we lack moral knowledge.

[22] We are only concerned with *robust* moral realism – the view that moral facts are (i) independent of minds, actual or idealized, and (ii) causally inert (Shafer-Landau 2003; Enoch 2011), as many debunkers think that robust moral realism is *particularly* susceptible to EDAs.

2. Evolution has strongly influenced our moral faculties in a way that defeats the justification of our moral beliefs.

3. There is no independent reason to accept the reliability of our moral faculties.

Therefore,

4. If moral facts are mind independent, then we lack moral knowledge.

Some critics object that EDAs overgeneralize in a costly or implausible way to non-moral domains like, say, the perceptual and epistemic (Shafer-Landau 2012; Vavova 2014; Das 2016: 432-33). If, for instance, EDAs undermine the justification of our beliefs in the epistemic domain, they are too costly or implausible – this overreach would undermine the justification for most, if not all, our beliefs. And this overreach by EDAs would be good reason to reject them, even in the moral domain.

In this paper, I argue that the overgeneralization objection is largely ineffective against the debunker, as it falls prey to a dilemma: either we have an independent reason to think that EDAs fail in a given domain, or we don't. If we have such a reason, then EDAs don't overgeneralize to that domain. But if we (i) lack an independent a reason, (ii) and evolution would distort beliefs in that domain, the evolutionary debunker needn't worry about EDAs overgeneralizing – it would be no more worrisome than debunking in the moral domain. Either way, the objection shouldn't worry the evolutionary debunker.

The paper proceeds as follows. First, I frame the debate over EDAs and review three formulations of the overgeneralization objection: analogy, entailment, and entanglement. Then I argue that each version of the overgeneralization objection faces the independent reason dilemma. Thus, at worst, EDAs overgeneralize benignly – under such conditions, debunking in a non-moral domain isn't any worse than debunking in the moral domain.

2 | The Overgeneralization Objection

The overgeneralization objection holds that EDAs overreach in a costly and implausible way: if the evolutionary origins of our moral faculties defeat the justification of our moral beliefs, they would also undercut the justification of beliefs in other domains like, say, our perceptual beliefs. However, such implications would be costly or implausible. This translates into a good reason to reject EDAs, even in the moral domain.

In this section, we frame EDAs as undercutting defeaters – then explain three ways to formulate the overgeneralization objection.

2.1 | Framing the Debate

The moral realist holds that many of our moral beliefs are justified by being grounded in mind-independent moral facts (Shafer-Landau 2003). The debunker, however, claims that our moral beliefs lack something required for justification like, say, lacking evidential sensitivity or an explanatory role, given the evolutionary origins of our moral cognition – these origins are a kind of undercutting defeater for our moral beliefs (Braddock 2017; Bogardus 2016: 637; Lutz 2018).

The moral realist, then, must either show that (a) our evolution isn't a defeater because, say, evolution doesn't fully explain our moral faculties (Machery and Mallon 2010; Cline 2015); (b) our evolution fails to defeat the justification of our moral beliefs because we can correct evolutionary influences using tools like, abstract reasoning and sophisticated methodologies (FitzPatrick 2015); or, (c) we have an independent reason to trust our moral faculties – a 'defeater-defeater' that blocks EDAs (Pollock 1986: 45-58).

Moving forward, we focus exclusively on (c). We frame EDAs as undercutting defeaters, where the justification of moral beliefs is undercut by their evolutionary origins – EDAs don't put the onus of proof on the moral realist. And given a plausible explanation of moral cognition that explains our moral beliefs without presupposing or entailing their truth, formulated as an undercutting defeater, we need an independent reason to trust the reliability of our moral faculties. First, though, we should clarify the independent reason objection.

2.2 | Three Versions of the Overgeneralization Objection

There are a few versions of the overgeneralization objection: analogy, entailment, and entanglement. We briefly distinguish these versions, then examine how the independent reason dilemma interacts with each version of the objection.

The *analogical* version (Shafer-Landau 2012: 23-5; Vavova 2014: 82-3) holds that EDAs applies to non-moral domains like, say, the perceptual domain, by analogy, with skeptical results – despite the reliability of our perceptual faculties. For instance, if EDAs apply to the perceptual domain, EDAs implausibly overreach, in that for one thing, they would undermine our knowledge that we evolved. And given this overreach, we should be suspicious of EDAs applied to moral beliefs. This overreach results because EDAs would not only apply to the moral domain, but also the epistemic domain as 'we would still take ourselves to have [epistemic] reasons, due to evolutionary forces' (Cline 2018: 155).

The *entailment* version holds that moral and epistemic reasons are the same relation in that each kind of reason *categorically warrants* beliefs and actions; and, if EDAs entail skepticism about moral reasons, then they also entail skepticism about epistemic reasons (Rowland 2016: 161-5; 2013). The difference between moral and epistemic facts is what they warrant: moral facts warrant action; epistemic facts warrant action. Epistemic and moral facts are relations with different relata: if EDAs defeat moral beliefs that entail moral reasons, then it would also defeat beliefs that entail epistemic reasons. But if this is right – and we have evolutionary reasons to doubt that there are moral reasons given their categorical nature – then we should be skeptical of realist epistemic reasons, too.

Finally, the *entanglement* version (Case 2018) holds that epistemic facts are entangled with moral facts. On this view, certain plausible epistemic assessments presuppose mind-independent moral reasons: if we reject epistemic reasons, then we must also reject these assessments. And if epistemic realism is more plausible than moral realism, then EDAs overgeneralize in a costly way. For example, if Bob's epistemic justification for the belief that *Sally is guilty of murder* is tied to how fastidiously he investigated her alleged crime

– the more fastidious he was, the more his suspicion is justified, ceteris paribus – then the epistemic facts are entangled with moral facts. So, EDAs are not confined to debunking beliefs in the moral domain *if* the epistemic and moral domains are entangled.

In later sections, we will examine each version of the overgeneralization objection; then I argue that the independent reason dilemma undermines each version.

3 | The Independent Reason Dilemma

3.1 | The Independent Reason Condition

It is widely held that EDAs have an independent reason condition: they succeed only if we have reason to think that evolution would be epistemically distorting; and, (b) we lack an independent reason to think that we have reliable access to realist moral facts (Joyce 2006: 216-18; Morton 2016: 8-9; Street 2006: 123-4; Vavova 2014: 81).

We should discuss what counts as an 'independent reason': a reason, *R*, is *independent* if either (*i*) R is outside the domain targeted by an EDA, or (*ii*) R is a constraint that applies to arguments and reasons across various domains. An example of the first: a successful IBE argument, for the reliability of our perception, would count as an independent reason to trust the reliability of our perceptual beliefs. As an example of the second: if assuming the unreliability of the faculties in a given domain like, say, the perceptual domain, would be self-defeating, that's an independent reason to take that faculty as reliable – we can't do otherwise in way that *isn't* self-defeating.

Also, EDAs would succeed *only if* there is no independent reason to think that we have reliable access to realist facts in that domain. If we have an independent reason to believe that we have reliable access to mind-independent facts in a particular domain, EDAs wouldn't overgeneralize to that domain – an independent reason blocks overgeneralizing. We find a similar point in the peer disagreement literature: some hold that we shouldn't evaluate our disagreement with an epistemic peer – someone who is equally informed and reliable about the dispute as us – by using reasoning which led to disagreement; we should appraise our disagreement with peers on neutral epistemic ground (Christensen 2011; Lackey 2013: 243; Vavova 2018).

The evolutionary debunker has good reason to posit an independent reason condition: if there were an independent reason, evolution wouldn't undermine the epistemic status of our moral faculties – we would have an independent reason to trust such faculties[23]. As Vavova (2014) observes:

> We can now see what the debunker thinks we need if we are to avoid her challenge: a reason to think that we are not mistaken in our evaluative beliefs that doesn't simply presuppose the truth of those beliefs. *This reason is, in some sense, independent of what is called into question*. This

[23] The moral realist might object that we cannot '… determine if we are likely to be mistaken about morality if we can make no assumptions at all about what morality is like […] but if we cannot take for granted any of our beliefs about the evaluative truths, then we cannot infer that they come apart from the adaptive beliefs' (Vavova 2014: 92).

But this objection isn't convincing *if* debunkers can reasonably assume that (a) the space of moral beliefs *vastly outstrips* the space of moral truths (cf. Wielenberg 2016: 510; Street 2006: 122), and (b) selection pressures would favor *moral* beliefs *for their content*, not their truth (i.e. selection cares what we believe, *not if we're right*). But (a) and (b) are plausible assumptions for all parties in the dispute to accept. By analogy, I need not know the winning lottery number to reasonably believe that it is unlikely that my ticket is a winner.

> explains why the debunker asks us to bracket our evaluative beliefs […] and to focus only on the origin story (81—emphasis mine).

Independent reasons aren't subject to *distorting* evolutionary processes in that such reasons cannot rely on either (a) intuitions about whether beliefs in a specific domain (in this case, the moral domain) are about mind-independent facts, or (b) a belief or intuition that *p*, where the truth of *p* would imply a mind-independent fact in a domain where we have good reason to expect that evolution would be *distorting*. For example, if there were a plausible IBE-style argument for moral realism – the view has greater explanatory power than its competitors, say – then that argument would be an independent reason to reject EDAs. For the sake of argument, we assume that the reasoning underlying IBE inferences is free of distorting evolutionary influences; and that such an argument would give us a reason to trust our moral faculties. (Similarly, we may have good explanationist reasons to reject external world skepticism – see McCain (2016)).

But the independent reason condition leads to a dilemma: either we have an independent reason to think that EDAs fail in a given domain, or we don't. If we such a reason, EDAs don't overgeneralize to that specific domain. But if we lack such a reason—and we have good reason to think that our beliefs in that domain were distorted by evolutionary processes—then EDAs overgeneralizing to that domain would be no more costly or implausible than debunking in the moral domain. Either way then, the objection shouldn't worry the evolutionary debunker.

Next, we examine versions of the objection given the independent reason dilemma.

3.2 | Overgeneralizing by Analogy

If EDAs debunks moral facts, they would apply *mutatis mutandis*, to our justification for our beliefs in realist facts in other domains. But, the objection goes, epistemic realism – the view that if fact F makes proposition *p* more probable, then F is a mind-independent reason to believe that *p*[24]; Street 2009: 219; Sylvan 2016: 365; Rowland 2013: 3 – is harder to deny than moral realism; it is plausible that epistemic reasons are realist. If EDAs debunk beliefs in the moral domain, then they would also *mutatis mutandis* debunk beliefs in other domains like, perception. As our beliefs in these non-moral domains are justified, this overreach by EDAs is costly and implausible – and good reason to reject EDAs, even when operating in the moral domain.

A | AN EXAMPLE FROM PERCEPTUAL REALISM

Consider the claim that evolution influenced our perceptual faculties. After all, to the extent that perceptual experiences influence behavior – and evolution is sensitive to our behavior — this is plausible. But this raises the question of whether we should trust our perceptual faculties given that evolutionary influenced them.

It is worth quoting Shafer-Landau (2012: 21) at length:

[24] Although metaethicists take epistemic reasons to be facts (Sylvan 2016: 365; Street 2009: 219), epistemologists take them to be mental states (Turri 2009). Since the debate over EDAs is on metaethical turf, I'll procedurally side with the metaethicists.

> As with morality, evolutionary pressures in the perceptual domain have as their confirmed doxastic effect the cultivation of adaptive dispositions. […] We can tell that dispositions to hold false perceptual beliefs are likely to be *maladaptive only if we can identify some false perceptual beliefs, show that they tend to undermine fitness and make inferences from those cases* […] If we are required to suspend judgment about all perceptual beliefs – *as we must, if required to do so in the moral case* – then we will most likely not be in a position to confirm the reliability of our perceptual faculties. (emphasis mine—also see Vavova 2014: 82-3; 92).

The worry is that if we run an EDA against our perceptual faculties, then we should be skeptical of our perceptual beliefs for evolutionary reasons. How *could* we test whether evolution perniciously influenced our perceptual faculties? We lack a better way to push back against EDAs applied to our perceptual faculties than the moral domain — given that this is untenable in the perceptual case, it is untenable in the moral case too. Thus, we should reject EDAs altogether.

There are a couple responses to this worry though.

As I said earlier, EDAs are typically framed as an undermining defeater. But evolution is defeating *only if* evolution in that domain would have been *distorting*. Debunkers hold that moral beliefs conducive to cooperation have been adaptive *despite their truth* like, say, the truth-value of the belief that *we are obligated to feed and nurture our children* is irrelevant where evolution is concerned (Joyce 2006: 131). This difference explains why evolution only sometimes defeats doxastic justification: they defeat the justification of such beliefs *only if evolution would have distorted beliefs in that domain*, and we lack an independent reason to think otherwise. Just because evolution influenced our perceptual faculties isn't sufficient to defeat the justification of our perceptual beliefs. And reliable perception would likely be adaptive (Carruthers 1992: 183-7; Fales 2009).

Some philosophers push back here: they hold that evolution wouldn't care about reliable perception. Evolution is indifferent to true beliefs in that survival doesn't require belief *at all*; and even then, it wouldn't require *true* beliefs – false beliefs can be adaptive too (Plantinga 2011: 328-9). The idea here is that while true beliefs are better guides to the world than false beliefs, ceteris paribus, this doesn't assuage such worries. If reproductive success could be secured as easily with false beliefs as true ones, then evolution would have obliged (Plantinga 2011; Stich 1993: 56-70). For instance, if Paul can avoid a tiger, then it is irrelevant that Paul falsely believes that tigers are friendly vegetarians; if Paul more strongly believes that tigers don't like to catch individuals they're chasing because, say – it's more fun for them – his false tiger beliefs won't get him killed.

We might initially think that evolution would favor individuals with reliable perceptual faculties – but we must still contend with Plantinga's skeptical reasons. While Plantinga is right that many creatures survive without beliefs *at all* like, say, bacteria, we could make a similar point about wings: many creatures survive without wings – evolution doesn't care about wings or flight *per se* – but for some creatures, there is adaptive value to wings that permit flight given their phylogenetic history. Similarly, action-directing beliefs 'once adopted, are adaptive if true and harmful if false' (Boudry and Vlerick 2014: 69; Deem 2018).

Also Plantinga's claim that evolution doesn't care about true beliefs in that compensating false beliefs could be as adaptive as true beliefs is suspicious. For one thing, not any false beliefs will do: rather, false beliefs can be adaptive *only if* coupled with the right sort of compensating false beliefs. Belief-forming processes would have to be able to distinguish compensating false beliefs from other false beliefs, which would be costly and hard. And, presumably, this would require a model of the world to allow such processes to identify false, but compensating beliefs – it would be easier to just form true beliefs to begin with, ceteris paribus.

So, not only do we lack a good reason to think evolution would distort perception, but we also have a candidate for an independent reason to trust our perception: explanatory virtues that favor the realist hypothesis in the debate over external world skepticism, where we can infer that the realist hypothesis is true, as it better explains our perceptual experience than the skeptical hypothesis (Vogel 2005; Huemer 2016). Even if abductivist strategies fail (Gifford 2013), they would still be a good candidate as an independent reason: an appeal to explanatory virtues, at least, seems to not be subject to evolutionary distortions (Vogel 2005), which in turn, satisfies the first horn of the independent reason dilemma. Next, we examine an example from epistemic realism.

B | EXAMPLE FROM EPISTEMIC REALISM

Epistemic realism can be formulated as the view that if fact F makes proposition *p* more likely to be true, then F is a mind-independent reason to believe that *p* (Street 2009: 219).

The fact that there are dinosaur bones in the ground, for instance, raises the probability that there were dinosaurs; and so, it is a mind-independent reason that warrants the belief that 'there are dinosaur bones in the ground' absent any countervailing factors (Rowland 2013: 3). Some moral realists hold that there is a parallel between epistemic reasons categorically warranting belief, and moral reasons categorically warranting action. Even if one rejects moral realism, many hold that epistemic realism is independently plausible. And epistemic reasons seem categorical just like moral reasons – they hold despite things like our aims and desires (Cuneo 2007: 117-22; Rowland 2013: 3-8).

Consider the first horn of the dilemma here: we have an independent reason to frame epistemic facts categorically. What would count as an independent reason? A plausible candidate is self-defeat: any EDA that, in overgeneralizing to the epistemic domain, implies that we have good epistemic reasons to reject epistemic reasons would be self-defeating. And if EDAs entail that there are no epistemic reasons to believe anything, then *ipso facto* they are toothless. EDAs overgeneralizing to the epistemic domain, in a self-defeating or toothless way, would be an independent reason to reject them (Cuneo 2007: 117-8; Kyriacou 2016).

A critic might object that we could reject epistemic realism, without rejecting epistemic reasons *simpliciter* – epistemic constructivism would be true (Street 2009). However, let's assume *arguendo* that rejecting epistemic realism collapses into epistemic nihilism (perhaps alternatives such as epistemic constructivism are unstable). If so, then EDAs are self-defeating or toothless. And this costly and implausible implication is good reason to reject them. Call this *the self-defeat or toothless objection.* We might formulate this as:

> If EDAs are self-defeating or toothless—as a result of overgeneralizing to the epistemic domain—then that is a good reason to reject EDAs (even in the moral domain). Call this *the defective claim.*

But the defective claim is but one approach to dealing with the self-defeating or toothless nature of EDAs overgeneralizing. There is a debunker-friendly approach: the problem is not that EDAs are self-defeating or toothless, and thus flawed – the correct takeaway, rather, is that specific domains have features that guarantee that EDAs will fail *in that domain* – but not because EDAs are defective otherwise like, say, when operating in the moral domain. It is, instead, because of relevant epistemic differences between different domains.

Here's an analogy: the fact that a heuristic works in one situation, but fails in a different situation, doesn't convict the heuristic in the former situation – it convicts applying the heuristic to the latter situation (the heuristic isn't the problem; wrongly applying it is). EDAs may work similarly: the epistemic domain has features that pose trouble for the charge that EDAs viciously overgeneralize, but not such that it tells us whether they work in the moral domain. There might be salient epistemic differences, between the epistemic and moral domains, in that EDAs only fail when applied to the epistemic domain. Here are a couple examples:

> **Asymmetrical Self-Defeat or Toothlessness**
>
> Arguments for *epistemic* nihilism are self-defeating or toothless, but arguments for *moral* nihilism aren't. We cannot, however, say this about arguments for moral nihilism – they aren't susceptible to self-defeat or toothlessness worries.

> **The Believability Gap**
>
> In some domains, we can believe true things; but not in others. For example, if moral error-theory is true, we can believe it without inconsistency. Whereas, if evolution distorted our perceptual faculties to such a degree that we cannot trust them, then we cannot *believe* that 'evolutionary has distorted our perceptual faculties to such a degree that we cannot trust them' – this would require believing that our perceptual faculties are reliable enough such that we can believe that evolution happened, given our empirical evidence.

This suggests a different way to think about EDAs overgeneralizing:

> If EDAs are self-defeating or toothless – as a result of overgeneralizing to the epistemic domain – then that is a good reason not to apply them to the epistemic domain; but it is not a good reason to reject them altogether. Call this *the applicability claim.*

The idea is that overgeneralizing that is costly or implausible like, say, it is self-defeating, is a good reason not to apply EDAs to certain domains – it is not a good reason to reject EDAs altogether. The moral skeptic can argue that self-defeat or toothlessness is a reason to adopt moral, rather than normative, skepticism. And debunkers can turn the tables on their realist critics here: the fact that EDAs are self-defeating or toothless in a particular domain could tell us whether EDAs are applicable to a given domain, but not whether they are defective overall. We can show this, by running the argument from the beginning of the paper, swapping out ‘moral’ with ‘epistemic’[25]:

> 1. If (a) **epistemic** facts are mind-independent, (b) evolution has strongly influenced our **epistemic** faculties in a way that defeats the justification of our **epistemic** beliefs, and (c) there is no independent confirmation of the reliability of our moral faculties, then we lack moral knowledge.
>
> 2. Evolution has strongly influenced our **epistemic** faculties in a way that defeats the justification of our **epistemic** beliefs.
>
> 3. There is no independent reason to accept the reliability of our **epistemic** faculties.
>
> Therefore,
> 4. If **epistemic** facts are mind independent, then we lack **epistemic** knowledge.

The moral realist might argue that this argument *sure looks sound*. However, we already know that we can’t accept it: if we accept (4), we cannot accept (2). If we can’t trust our epistemic faculties, then how would we know that evolution occurred? The same worry, however, doesn’t apply to the argument if we swap out ‘epistemic’ for ‘moral’ – while this argument might be false, it wouldn’t self-defeat. The idea is that if the moral realist holds that EDAs overgeneralize to the epistemic domain in a self-defeating or toothless way – and this is good reason to reject EDAs – then they must show that the lesson of overgeneralizing in an implausible or costly manner is good reason to reject EDAs, not

[25] This could also be done replacing ‘moral’ with ‘perceptual’ in the argument.

that it implicates *how the EDA was applied*. The debunker can rightly reply that this is a plausible interpretation of overgeneralizing that doesn't entail we should reject EDAs altogether (even in the moral domain). And thus this *tu quoque* challenge to our epistemic faculties *cannot* be sound, despite appearances, as it is self-defeating – whereas the same argument, with 'moral' instead of 'epistemic' swapped out, isn't.

We might also worry that if EDAs apply to the moral and epistemic domains, by analogy, this would indicate that there is something wrong with EDAs, even when applied to the moral domain. This move, however, assumes that the moral and epistemic domains are similar enough, in epistemically relevant ways, such that EDAs would apply, by analogy, to both domains. If the applicability claim is right, then EDAs wouldn't overgeneralize to the epistemic domain (but would to the moral domain) in that they have salient epistemic differences.

3.3 | Overgeneralizing by Entailment

Consider overgeneralizing by entailment: moral and epistemic reasons are categorical. If we reject moral reasons because they are categorical, we must reject epistemic reasons too. This should worry the debunker only if rejecting epistemic realism would be costly or implausible in, say, a self-defeating or toothless manner. Rowland explains:

> [Once] we know that there are instances of the basic relation R, the weirdness of the properties that Rs have is eliminated as a sufficient reason to be sceptical of Rs *even though we only know that there are a subset of Rs*. Even if the features of Rs that are metaphysically weird still provide

> a sufficient reason to doubt that there are things other than Rs that have these features (2016: 167—emphasis mine).

If our evolution gives us reason to doubt that there are realist moral reasons, *and the entailment approach is right*, then EDAs would be self-defeating: EDAs would give us reason to both accept and reject the conclusion of the argument (Streumer 2017: 170-72; Cuneo 2007: 117-18). The argument relies on an evolutionary reason to reject realist oughts, but seems to require such oughts to motivate the argument. The point isn't that a self-defeating argument works, *even on the entailment approach*; rather, the point is that the entailment approach could cause trouble by costly or implausible overgeneralizing – if only realist epistemic reasons can justify beliefs (Kyriacou 2016).

Thus far we haven't discussed how to individuate domains. For our purposes, I assumed that the kind of fact individuates domains like, say, the moral and prudential domains are distinct because they would contain different kinds of normative facts. If, however, moral and epistemic reasons are 'fundamentally the same relation,' moral and epistemic reasons occupy the same domain.

Here the debunker may worry: if she wants to hold epistemic realism, *but reject moral realism*, her debunking argument may wrongly overgeneralize. If her EDA undermines her moral beliefs, then it undermines all her beliefs[26]: taking epistemic and moral reasons to be the same kind of reason serves to raise the threshold for denying moral realism. The

[26] We are, of course, assuming for the sake of argument that without categorical epistemic reasons, we would be without epistemic reasons altogether (but it's not quite that simple).

strategy, for the critic, is to keep the debunker from accepting epistemic realism, while also rejecting moral realism – she must be forced to accept or reject both if moral and epistemic reasons are fundamentally the same kind of reasons with different relata.

The entailment version, however, is vulnerable to the independent reason dilemma. And notice that *if* the moral and epistemic domain cannot be distinguished, in that reasons in each domain *both categorically warrant* beliefs (the epistemic domain) and action (the moral domain), then we should treat such reasons as if they belong to the same domain, *for the purposes of evolutionary debunking* (if the evolutionary challenge to such reasons targets their categorical nature). But if this normative taxonomy is correct, and we should expect that evolution would be distorting in a hybrid domain (Street 2009: 235-6), then without an independent reason to believe that we have epistemic access to realist moral *and epistemic* reasons, we should reject moral and epistemic realism.

And if we have an independent reason to believe that we have epistemic access to realist moral and epistemic reasons (taken together), then EDAs cannot overgeneralize. This is where the charge that EDAs operating in the epistemic (and perceptual) domain is self-defeating or toothless is salient. If such charges could be made to stick—even though, as we argued earlier, this is questionable—then such charges would count as an independent reason to think that we *shouldn't* collapse the moral and epistemic domains – EDAs fail in the latter domain, but not the former (as collapsing the domains results in self-defeat or toothlessness).

We can plausibly argue that the fact that EDAs would be self-defeating or toothless, if we collapsed the moral and epistemic domains into a hybrid domain, is good reason for the *moral* skeptic to reject the claim that moral and epistemic reasons belong in the same domain. For example, if arguments for moral nihilism aren't self-defeating or toothless, but arguments for epistemic nihilism are either self-defeating or toothless[27], that might be good reason to distinguish we between moral and epistemic reasons, and not collapse the moral and epistemic domains into a hybrid domain. We might even grant that moral and epistemic reasons are fundamentally the same kind of normative reasons, but hold that there are other features of moral and epistemic reasons that are good reason to distinguish between the moral and epistemic domains.

A critic might reply that the point above doesn't save EDAs from the overgeneralization charge: moral and epistemic reasons could still belong to the same domain. And if so, then EDAs could overgeneralize in a costly or implausible manner. However, this misses the point of the response above: it is defensive in that the entailment approach holds that moral and epistemic reasons belong in the same domain; in that case, EDAs either don't apply, or they are self-defeating or toothless when applied outside the moral domain. But a moral skeptic could plausibly respond here that we haven't been given good reason to collapse the domains, other than the features in common. The debunker could accept this similarity, and reply that self-defeat or toothlessness would be good enough reason not to collapse the moral and epistemic domains into one.

[27] Streumer (2013) is a normative error-theorist who argues that epistemic nihilism is true, but we cannot believe that it's true; and he argues that this is a feature, not a bug, in that it allows us to escape charges that epistemic nihilism is self-defeating or toothless.

On the other hand, if we lack an independent reason to think there are categorical moral and epistemic reasons – like, say, if overgeneralizing didn't result in self-defeat – then overgeneralizing would be benign: it would be no more worrying than debunking in the moral domain. If there are equally good reasons to be skeptical of *epistemic and moral reasons*[28], the fact that (i) the moral and epistemic domains form a hybrid domain, and (ii) we lack an independent reason to trust beliefs in that domain, would mean that EDAs apply benignly to the epistemic domain.

3.4 | Overgeneralizing By Entanglement

The entanglement version holds that if epistemic realism is true, and some moral and epistemic reasons are appropriately entangled, then epistemic and moral realism stand or fall together, with regard to their justification. Case (2018) argues that,

> [The] epistemic standing of a belief depends on how diligent an inquirer the agent has been; this in turn depends on facts about *the moral context of the agent's situation*. Hence moral and epistemic facts are intertwined […] a decision not to seek additional information constitutes investigative negligence depends on the practical, and specifically *moral*, context in which the decision is being made (1 – original emphasis).

And,

[28] If rejecting epistemic facts collapses into epistemic nihilism.

> [The] epistemic status of an agent‘s belief depends on what evidence the agent possesses when the assessment is being made. Although evidence possession is synchronic, *ethical considerations are relevant to whether or not evidence is possessed by an agent at a time*. So moral facts are entangled with epistemic facts (1 – emphasis mine).

The entanglement version rests on the following claims:

> (a) We have compelling reason to take epistemic realism seriously; and,
>
> (b) Epistemic facts are appropriately entangled with moral facts.

The entanglement approach holds that some epistemic facts hold *only if* certain moral facts do; and we have reason to accept realist epistemic facts[29] — call this *entangled epistemic realism*. For example, Joe believes that Mary cheated on him; and his epistemic justification for this belief depends on how thoroughly he investigated the matter: if he merely found weak evidence for her infidelity, and failed to investigate further, then his epistemic reasons for believing that Mary was unfaithful are weak. But *fastidiousness* is a moral notion (in a board sense); so, the epistemic facts (e.g. whether Joe is epistemically justified in believing that Mary cheated) are entangled with moral facts (e.g. the degree to which Joe was thorough in his inquiry). Thus, there are cases where the epistemic and moral facts are entangled.

[29] The entanglement approach is a parity argument. And parity arguments are *bi-directional* – for instance if epistemic and moral facts are entangled, then we must either accept or reject both. Depending on which way we go, it may be bad for the moral realist (if, say, rejecting epistemic realism isn’t a big deal).

Case recognizes, however, that (a) and (b) alone don't vindicate moral realism. After all, it could be that epistemic realism and moral constructivism are true, and epistemic facts are entangled with constructed moral facts. He thinks that the following (independently plausible) principle will avoid this issue:

> If the salient moral facts are less robustly realist than the salient epistemic facts, then epistemic facts would be hostage to desire. Call this *the robustness principle*.

Even if we grant the robustness principle, however, it doesn't do what Case needs. Let's distinguish entangled and non-entangled realist epistemic facts: the former are epistemic facts entangled with moral facts like, say, the fact that Joe is justified believing that *Mary cheated on him* is entangled with how fastidious he was in his inquiry. And the latter are epistemic facts wholly separate from moral facts like, say, the fact that dinosaur fossils were found is reason to believe that *dinosaurs roamed the Earth*. The intuitive appeal of epistemic realism may be exhausted by non-entangled realist epistemic facts – where we would have reason to accept epistemic realism, but still reject the claim that some moral and epistemic facts are entangled. Even given the appeal of epistemic realism, we need a good reason to hold that some epistemic facts are entangled with moral facts such that they are not 'hostage to desire', if overgeneralizing by entanglement is to work.

And, more to the point, the viability of entangled epistemic realism depends on whether there is an independent reason to think there are categorical moral facts – otherwise, the debunker could hold that some epistemic facts would be 'hostage to desire,' without having to admit that epistemic facts *simpliciter* are hostage – and the debunker shouldn't find this more worrying than debunking in the moral domain. If the moral realist wants the entanglement approach to push overgeneralizing in a vicious way, then she needs an independent reason to think that there are realist moral facts.

The moral realist might appeal to her intuition that there are epistemic facts not 'hostage to desire'; and if the entanglement approach is right, then we have epistemic access to *some* categorical moral facts. Here debunkers might reply that there is reason to worry that the appeal of epistemic realism, when epistemic facts are entangled with moral facts, might be suspect given prior reason to worry that evolution would be distorting in the moral domain. And debunkers have a plausible account of why epistemic facts seem to *categorically warrant* belief: if beliefs influence behavior, taking facts to categorically warrant beliefs may have been adaptive (Street 2009: 235-6).

If there *isn't* an independent reason, the best that the entanglement approach can establish is that epistemic facts are *somehow* entangled with moral facts *of some kind* — and this wouldn't satisfy the moral realist in that the moral anti-realist could accept this too. The debunker might argue that if the entanglement approach is right – some epistemic facts are entangled with moral facts – then if there is reason to worry that evolution would be distorting in the moral domain, the entanglement approach would be reason to be skeptical of entangled epistemic facts too. We should be skeptical that there are entangled epistemic facts without an independent reason to accept moral realism. And then, at most, EDAs would benignly overgeneralize.

4 | Conclusion

In this paper, I argued that the overgeneralization objection faces the independent reason dilemma: either we have an independent reason to think that we have epistemic access to mind-independent facts in a given domain, or we do not. If we have such a reason, then EDAs fail to overgeneralize. If we lack such a reason, but have good reason to think that evolution would have been distorting in that domain, EDAs benignly overgeneralizes. Debunkers shouldn't worry that EDAs overgeneralize in a costly or implausible way – it either doesn't overgeneralize, or it overgeneralizes benignly.

Chapter 4: Moral Fictionalism and Moral Risk

1 | Introduction

Many philosophers believe moral skepticism or error-theory because of, say, intractable disagreement and evolutionary influences on our moral cognition. And yet, despite their metaethical views, many of them still feel the pull of moral concerns as much as they did prior to adopting skeptical or error-theoretical views. This should prompt a question: are there are good reasons to keep believing, thinking, and acting morally, despite holding such metaethical views?

There are three prominent answers to this question in the literature[30]. Moral *eliminativists* hold that it would be better overall if we forgo moral belief, thought, and action totally. Moral *fictionalists* hold that we should discard our moral beliefs, but keep thinking and acting in moral terms *as a robust pretense.* Finally, moral *conservationists* hold that we should keep our moral beliefs, thoughts, and actions – we should carry on like before.

In this paper, I defend moral conservationism: evidence from cognitive science says that we cannot help but keep our moral beliefs. And, acting on our moral beliefs allows us to mitigate moral risk – when thinking and acting, we could be wrong about weighty moral

[30] There are other views, too, like moral revisionism (Cline 2018), and negotiationism (Eriksson and Olson 2019) – but we will focus solely, in this paper, on the three views mentioned in the introduction.

matters, and sometimes act in seriously wrong ways (call this *the moral risk approach*). We also have defeasible reason to reject moral eliminativism in that it would likely (a) deprive us of practical goods like, long-term cooperation; and, (b) exacerbate moral risk – the risk that our moral thinking is wrong, and we would do something seriously wrong if we acted on such thinking. Securing practical goods like, say long-term cooperation, and mitigating moral risk, favor moral conservationism over its rivals.

Here's an outline of the paper. First, I introduce moral fictionalism, and explain its two rivals: moral eliminativism and moral conservationism. Then we discuss how the moral fictionalism approach purports to secure practical goods of thinking and acting morally. Moral fictionalism, however, has an implausible view of how to discard, and avoid forming, moral beliefs – something we must do to practice moral fictionalism. Finally, I argue that mitigating moral risk is best done by keeping our moral beliefs. The weight of such reasons favors moral conservationism over its rivals.

A PRELIMINARY OBJECTION

A critic might worry that appealing to moral risk to generate moral oughts (to, say, justify avoiding moral risk) and enable moral thinking and action, begs the question against the moral skeptic and error-theorist (Joyce 2005: 288). But this objection misunderstands that the moral risk approach is motivated by *practical* reasons like, enhanced self-control – these aren't *moral* reasons. This approach should appeal to individuals who want to live a moral life, and avoid serious moral risk – reasons that can be framed in broadly Humean terms, as a function of one's desires.

2 | Moral Fictionalism

Moral fictionalism holds that we should *act as if* moral reasons hold to secure *practical* goods like long-term cooperation, but forego moral pretense in critical contexts like, say, a philosophy seminar, where such second-order beliefs are salient (Joyce 2001; 2005).

2.1 | Two Alternatives to Moral Fictionalism

Consider a couple of alternatives to moral fictionalism: moral eliminativism holds that we should forgo moral thought and action altogether, as this is better overall (Garner 2006). And moral conservationism holds that we should keep our moral beliefs, thoughts, and actions (Olson 2011). We review each view in turn.

A | MORAL ELIMINATIVISM

Moral eliminativism holds that we should abstain from morality altogether. This view is plausible enough: if beliefs in a given domain are systemically unjustified or false, they should be rejected, ceteris paribus, like say, atheists who discard their religious beliefs. Garner (2007) argues that there are benefits to abstaining from moral thinking and action altogether like:

1. Moral thinking and action will 'stabilize […] differential advantages the various parties initially have' (Mackie 1980: 154). Those who benefit from the status quo like, say in terms of wealth, could appeal to moral reasons to justify their status; they couldn't do this on moral eliminativism. Call this *the differential objection.*

2. On error-theory, there are no moral truths to settle a dispute as every ‘possible moral value and argument can be met by an equal and opposing value or argument’ (Garner 2007: 502). If there are no moral facts, but we retain moral discourse and practice, irresolvable disagreement would result – but not if we forgo moral thinking and action, and settle disputes by pointing to, say, prudential reasons. Call this *the swamping objection.*

Such objections don’t exhaust the case for eliminativism; but it would still be instructive to explain why they don’t work.

First, the differential objection isn’t reason to think that eliminating morality would be a good thing overall: moral reasons could be used by the rich to justify the status quo like, say, Bob arguing that his wealth was earned, and shouldn’t be taxed harshly; but moral reasons have also been used by reformers, like a civil rights leaders, who appeal to moral ideals in criticizing unfair advantages had by the status quo – moral eliminativists need to show that moral thinking and action does more harm than good *overall*, for those who are disenfranchised, for the differential objection to have any bite.

Second, *pace* the swamping objection: it isn’t obvious there would be more intractable disputes without moral truths, than with them, even though this *is possible*. For one thing, moral agreement could be adaptive by, say, facilitating long-term cooperation – there is evidence that moral disagreement is a costly signal (Kogelmann and Wallace 2018). It is unclear how many intractable disputes would result from eliminating or preserving moral thinking and action. And, without moral facts, we might expect more intractable disputes, as there would be no moral facts to help settle moral disputes. This is the idea behind arguments from disagreement against moral realism: moral disagreement should be expected in the absence of moral facts (Enoch 2009).

And without moral thinking and action, life would likely be terrible; cooperation would be nearly impossible – call this *the Hobbesian presumption*. Evolutionary debunkers point to the benefits of thinking and acting morally to explain why a moral sense evolved; and given that moral thinking action is costly and ubiquitous, it is probably adaptive. Morality facilitates long-term cooperation – and since long-term cooperative has practical value, morality does too[31]. And, on the moral eliminativist view, there is a very serious risk that whatever system replaces morality wouldn't facilitate long-term cooperation to the same degree as moral thinking and action. While there might be other goods had by eliminating morality, it is unclear if they are enough overall to justify doing so.

Finally, moral eliminativists also face the epistemic possibility of metaethical error for a couple of reasons.

First, we sometimes disagree with our past selves about metaethical matters. Sally has thought about Mackie's argument from moral disagreement, for instance – she is familiar with the argument, objections and replies, and so on. And yet, she accepts that Mackie's argument: intractable moral disagreement is best explained by the absence of moral facts – moral disagreement would be expected if there are no moral facts. But later, Sally finds

[31] Especially given our ability to cooperate over the long-term, in sophisticated ways, rather than our raw intelligence, explains our phenomenal success as a species (Henrich 2015).

Mackie’s argument weak, and the objections more convincing. She then moderates her metaethical view in response. If past and present Sally are equally diligent about such matters, their disagreement is evidence that (at least) one of them is wrong.

As Moller (2011) notes,

> [The] main reason for supposing there is a non-negligible possibility of error [is] … the subject matter involved is the sort of thing it is all too easy for people like us to be mistaken about; abstruse moral reasoning involving far-out cases and complex principles is something we find very difficult and are disposed to get wrong reasonably often (432).

There is a serious possibility that we have false metaethical views, in that we know that reflecting on the past reveals ‘the gaping holes in my prior ruminations. Things looked great back then; years later they don’t look nearly as convincing’ (Frances 2016: 289).

Second, there is peer disagreement: there are lots of people who are *prima facie* as smart, informed, and diligent as we are, but with whom we disagree (call them *epistemic peers*). This is a reason to take seriously the non-negligible possibility that we are wrong in our metaethical views. The same reasoning holds for moral skeptics and error theorists: they have epistemic peers who disagree over metaethical issues (Wedgwood 2014). This is some evidence that each party to the dispute could be wrong – assuming their views don’t exhaust logic space.

A critic might object that steadfasters reject peer disagreement as a reason to moderate their controversial beliefs (Kelly 2010; van Inwagen 2010), *pace* conciliationists, who hold that peer disagreement is a reason to moderate our credences (Christensen 2007), and equal-weighters, who hold that we should assign equal credence to our beliefs, and our peers' salient beliefs (Bogardus 2009). Steadfasters reject that peer disagreement is a reason to adjust our beliefs; but if so, then they wouldn't have a reason to act differently given the possibility of error. But that isn't a problem: steadfastness doesn't preclude the serious possibility of metaethical error – and that is all we need for the serious possibility of metaethical error to open the door to moral risk: if the moral eliminativist is wrong in her metaethical views, and acts as if there are no moral reasons, she runs the serious risk of moral wrongdoing. But the same cannot be said for the moral conservationist: even if she holds false metaethical views, yet she retains her moral beliefs, and continues to think and act in moral terms, then that the moral risk is mitigated for her (we discuss this more in later sections).

B | MORAL CONSERVATIONISM

Moral conservationism holds that we should 'think moralized thoughts' and that '[moral] belief is to be embraced rather than resisted' (Olson 2011: 198). Moral beliefs help to secure the practical goods of moral thinking and action. But *pace* moral conservationism, *if* a class of beliefs are systemically unjustified or false, then we would have *prima facie* reason to discard them like, an atheist would discard her theistic beliefs. But that said, moral conservationism has strengths over moral fictionalism: the motivational guidance of moral beliefs would, presumably, enhance self-control, and long-term cooperation.

Against this, however, Joyce has a couple of objections:

(i) Moral skeptics and error-theorists would find it hard to square their moral beliefs and metaethical views in that ‘even if they *could* somehow bring themselves to sincerely “forget” that they ever read Mackie’s book […] surely to embark on such a course is likely to bring negative consequences’ (Joyce 2005: 298).

And,

(ii) Holding false beliefs would be cognitively onerous in that ‘seemingly useful false belief […] will require all manner of compensating false beliefs to make it fit with what else one knows’ (Joyce 2001: 178).

Response to (i): the moral conservationist position entails holding conflicting beliefs – and is thereby susceptible to the charge of (doxastic; epistemic) irrationality. It is unclear, though, how worrying this charge is; it is *a* consideration against moral conservationism, but by itself, it’s not clear that it would be good reason to reject the view, unless ‘among the ends that an error theorist cares *most strongly* about [is] holding true and consistent beliefs and avoiding false and inconsistent ones … to realize these ends, the error theorist will have to jettison moral thought and discourse’ (Eriksson and Olson 2019: 128 – original emphasis). Our discussion will assume that the moral skeptic and error-theorist doesn’t hold take doxastic consistency as an overriding priority.

Response to (ii): it would be cognitively onerous to have false empirical beliefs that need compensating false beliefs to be useful; however, it isn't clear how that the issue of compensating false beliefs carries over to the *moral* domain. For instance, Joanne *falsely* believes that *everyone is morally equal*; but it is unclear why she would need false moral beliefs to compensate for her *false* moral equality belief. The moral fictionalist must explain why we would need false compensating beliefs *in the moral domain.* We might need them in the empirical domain in that if we had false beliefs, at least where they influence our decision-making, we would need compensating false beliefs. For example, if Paul falsely believed that tigers are friendly, to avoid being eaten, Paul would need compensating false beliefs like, say, tigers don't like to catch whoever they're chasing. It is unclear, however, that we need compensating false moral beliefs. And the moral is likely autonomous from the non-moral and empirical – moral conclusions don't follow from non-moral or empirical premises alone.

2.2 | The Purported Benefits of Moral Fictionalism

Moral fictionalists accept that moral pretense need not secure *all* the benefits of moral *belief*; it need only do a better job than the competing approaches. They also hold that taking a fictionalist stance toward morality (a) enhances self-control, and (b) facilitates long-term cooperation (without moral belief).

A | ENHANCED SELF-CONTROL

Joyce (2001) argues that moral pretense enhances self-control:

> Moral thinking, I contend, is […] an expedient [that] functions to bolster self-control against such practical irrationality. If a person believes that Φing to be required by an authority from which she cannot escape, if she believes that in not Φing she will not merely frustrate herself, but will become reprehensible and deserving of disapprobation—then she is more likely to perform the action. In this manner, moral beliefs can help us act in an instrumentally rational manner (184; cf. also Joyce 2006: 111).

Evolution would have favored individuals with a moral faculty like ours: a robust moral sense to serve as a psychological bulwark against factors that would otherwise confound our ability to act prudentially. If we preserve moral thinking and action, by engaging in moral pretense, then we would have practical reason to engage in moral pretense.

Someone might object that belief can be motivating, while pretense need not be like, say, a belief that I will be attacked by a tiger is terrifying; pretending likewise isn't. Joyce replies by arguing that moral pretense would motivate us to act morally (in the absence of belief). For instance, we might find that 'reading *Anna Karenina* may encourage a person to abandon a doomed love affair' or that 'watching *The Blair Witch Project* may lead one to cancel the planned camping trip in the woods' (2005: 303). Just as we might *actually* be afraid during the showing of a horror movie, moral pretense would help us think and act morally, enhancing our ability to refrain from imprudent actions.

B | LONG-TERM COOPERATION

Moral fictionalists hold that moral pretense can also facilitate long-term cooperation: if we *act as if* moral reasons have weight in our deliberations (except in critical contexts), then we will secure many similar benefits of moral thinking and action without having moral beliefs. As Joyce (2006) argues:

> The benefits that may come from cooperation—enhanced reputation, for example—are typically long-term values, and merely to be aware of and desire these long term advantages does not guarantee that the goal will be effectively pursued, any more than the firm desire to live a long life guarantees that a person will give up fatty foods […] If a person believes an action to be required by an authority from which he cannot escape, if he believes that in not performing it he will not merely frustrate himself, but will become reprehensible and deserving of disapprobation—then he is more likely to perform the action (111—emphasis mine).

(Joyce is talking here about *beliefs* – but he thinks the same holds of moral pretense). This benefit of morality comports with the reasons that evolutionary debunkers use to argue that a moral sense is adaptive (cf. Joyce 2006; Street 2006). Morality facilitates long-term cooperation in that, say, we would more likely cooperate with those who have reputation for fairness than those who don't.

2.3 | Problems with Moral Fictionalism

Moral fictionalists claim that moral pretense can help us secure many of the practical goods of thinking and acting morally, without belief; moral conservationists disagree. The problem is, in part, that moral fictionalists aren't clear about the notion of belief that informs their view. And it is plausible that the robust moral pretense, they need for their view, fosters moral *belief*. Let's review some troubles facing moral fictionalism.

A | THE DISCARDING PROBLEM

Moral fictionalists hold that we should discard our moral beliefs as (i) there is serious disvalue to *false* beliefs, and (ii) having *false* beliefs is cognitively onerous – they deny that we need moral beliefs to serve as a motivating guide to thinking and acting morally.

However, there are good reasons (philosophical, empirical) to think that forming beliefs is easy, and discarding them is difficult (cf. Mandelbaum and Quilty-Dunn 2015: 44-47; Alston 1988: 260-63). First, there is the issue of doxastic voluntarism: we lack direct control (forming, discarding) over our beliefs. It is not that we lack control over our beliefs *altogether*, but instead that the process of discarding beliefs is very hard like, say, Sally to discarding her belief that *Obama is the first black President of the United States*. If we could just discard our beliefs, then I must have 'effective voluntary control over whether I do or do not believe that the tree has leaves on it when I see a tree with leaves on it […] in broad daylight with my eyesight working perfectly' (Alston 1988: 264).

Second, there is experimental evidence that subjects under a cognitive load, like having to count backward from '100' in increments of '7', have a hard time discarding beliefs, but acquire them with ease. We should keep in mind that something like, say, deliberately not attending to something in one's visual field can induce a cognitive load (Mandelbaum and Quilty-Dunn 2015: 47). The rub for the moral fictionalist is that (i) we have a robust moral sense like, say, having the strong intuition that some actions are morally wrong; and (ii) not attending to our moral intuitions is enough to induce a cognitive load, making it very hard to discard beliefs – in just consciously suppressing our robust moral sense, we would plausibly have a hard time discarding our moral beliefs. This should trouble the moral fictionalist as situations where moral reasons are salient almost always involve cognitive loads induced by deciding what we ought to do morally.

However if we retain our moral beliefs, and still secure the practical benefits of moral thought and action, then we have a skeptical challenge to moral fictionalism: if we retain our moral beliefs, then it is a serious epistemic possibility that our moral beliefs, not moral pretense, which is a motivational guide to moral thinking and action – the moral fictionalist cannot rule out this epistemic possibility. So, difficulty in discarding beliefs should make us skeptical that we can secure such practical goods in the absence of *belief*.

B | THE SLIPPING PROBLEM

If we have robust enough moral dispositions, to engage in moral pretense, or to represent the world *as if* moral claims are true (and act accordingly), it is plausible that we have moral beliefs. But then it is unclear how we could morally pretend, to secure the practical goods of moral thinking and action, without forming moral beliefs. Joyce holds that we need robust moral dispositions to maintain an effective moral pretense – without them, we would need to engage in,

> […] an ongoing calculation that one makes over and over. It is not being suggested that someone enters a shop, is tempted to steal, decides to adopt morality as a fiction, and doing so bolster her prudent though faltering decision not to steal. Rather, the resolution to accept the moral point of view is something that occurred in the person’s past, and is now an accustomed way of thinking (2005: 306).

Joyce’s model of *moral pretense* is analogous to how a police officer might approach an undercover assignment: thinking and acting like a criminal becomes second nature to her. Consciously staying in character would be cognitively onerous for the police officer; and she would have to rely on robust dispositions to do her job. Likewise, a practicing moral fictionalist must be *robustly disposed* to engage in moral pretense.

But there are a couple of problems here for the moral fictionalist.

First, just as staying in character would be psychologically onerous for undercover cops like, say causing stress and losing one’s sense of self, it would be cognitively onerous to ‘stay in character’ with regard to moral reasons. The stressors here include the practicing moral fictionalist suppressing her doxastic inclinations like, say, Sally’s strong inclination to believe that ‘torturing babies is morally wrong.’ This would require lots of cognitive resources.

Second, if we are robustly disposed to reason and act as if moral reasons hold, then it is plausible that we have moral *belief* — or mental states that look an awful lot like belief. For one thing, robust moral dispositions and representations have the direction of fit of beliefs (i.e. p is a belief if it depicts the world as being in a state of affairs such that *p*). This way of distinguishing belief and desire has its critics (Sobel and Copp 2001), but it is sufficiently plausible for us. We only need reasonable doubt that the practicing moral fictionalist can think and act morally without using or forming beliefs. It is hard to see how a practicing moral fictionalist could rely on robust moral dispositions which don't *count as beliefs.* This worry is especially salient if we have an evolved moral faculty that lends itself to producing moral beliefs (Joyce 2006; Street 2006; Dwyer et. al. 2010). (We discuss this more in the next section on moral robustness).

The worry about robust moral dispositions *as* belief sits nicely with empirical evidence on the nature of belief (Mandelbaum 2014; Mandelbaum and Quilty-Dunn 2015): we cannot *entertain propositions without believing them* to some degree. This conflicts with the Cartesian idea that we take on beliefs in a deliberate, rational way – where we weigh the reasons for and against a proposition before believing it. Researchers found that when they put subjects under a cognitive load (by having them track numbers on a screen), then presented them with sentences, about crime statistics, flagged as true (in black) or false (in red), the subjects couldn't remain neutral. Research subjects read sentences about crime statistics while under a cognitive load; control group members didn't. The subjects were more likely than control group members to make inferences based on red (false) sentences like, say, recommending longer prison terms.

It is plausible that an important feature of belief is not just that they involve accepting a proposition, but they are also serve as the basis for action, and inferences in our practical deliberations – given this, research subjects plausibly *believed* the propositions expressed by the red (false) sentences to a greater degree than control group. This evidence strongly suggests that mental states that look a lot like belief are far easier to acquire than moral fictionalists suppose.

The slipping problem is to reconcile the requirement that we have to use robust moral dispositions to sustainably take a fictionalist stance toward morality with the requirement that we forgo moral belief. If robust moral dispositions to engage in moral pretense count as *beliefs*, we have an explanation for the motivational guide that allows thinking and acting morally, namely: moral belief. Given that moral beliefs are a good motivational guide – and it is unclear whether moral pretense could be such a guide, without turning into belief – we should be hesitant to adopt moral fictionalism. This point is underscored by our robust moral sense.

C | A ROBUST MORAL SENSE

The difficulty discarding moral beliefs is exacerbated by our robust moral sense: most people have a strong moral sense – they hold deeply ingrained moral values like, fairness and equality; for instance, people punish cheaters, even if its costly[32] – and it is difficult for them to strongly believe anything contrary to this robust moral sense like, say, that 'we should maximize pain, and minimize pleasure.' There is some evidence for this robust moral sense like, agreement across cultures on basic moral *values*, but not beliefs (Haidt and Craig 2004; Sauer forthcoming). Here are a couple examples:

(i) Basic evaluative attitudes

Selective pressures have bequeathed us a set of hardwired basic evaluative dispositions. Street (2006) lists some of them:

[32] See Boyd, Gintis, and Bowles (2010)

> 1. The fact that something would promote one's survival is a reason in favor of it.
>
> 2. The fact that something would promote the interests of a family member is a reason to do it.
>
> 3. We have greater obligations to help our own children than we do to help complete strangers.
>
> 4. The fact that someone has treated one well is a reason to treat that person well in return.
>
> 5. The fact that someone is altruistic is a reason to admire, praise, and reward him or her.
>
> 6. The fact that someone has been done one deliberate harm is a reason to seek his or her punishment (115).

People across cultures agree on claims like pain is bad, cheaters should be punished. And it is hard to find anyone who seriously disagrees with this list – much less, someone with very different beliefs about such matters like, say, the belief that pain is good, or we should reward cheaters.

(ii) Cross-cultural values

Cross-cultural data indicates that nearly everyone all over the world holds the same basic moral values, even though they emphasize them differently (Alfonso 2016; Schwartz 1994: 22):

> *Power:* Social status and prestige, control or dominance over people and resources
>
> *Achievement:* Personal success through demonstrating competence according to social standards.
>
> *Hedonism:* Pleasure and sensuous gratification for oneself.
>
> *Stimulation:* Excitement, novelty, and challenge in life.
>
> *Self-direction:* Independent thought and action-choosing, creating, exploring.
>
> *Universalism:* Understanding, appreciation, tolerance, and protection for the welfare of all people and for nature.

> *Benevolence:* Preservation and enhancement of the welfare of people with whom one is in frequent personal contact.
>
> *Tradition:* Respect, commitment, and acceptance of the customs and ideas that traditional culture or religion provide.
>
> *Conformity:* Restraint of actions, inclinations and impulses likely to upset or harm others and violate social expectations or norms.
>
> *Security:* Safety, harmony, and stability of society, of relationships, and of self.

This is strong, though not conclusive, evidence for a shared, robust moral sense that influences our moral thinking. We wouldn't expect agreement on values, across cultures, in the absence of a robust moral sense, given the universe of *possible* moral values – compared to, say, theories on the origin of the universe that have varied widely across cultures. This robust moral sense will be salient in the section on moral risk.

D | THE UNCERTAINTY PROBLEM

A primary contention by moral fictionalists is that robust moral dispositions allow the practicing moral fictionalists to gain many of the benefits of engaging in moral thinking and action, without relying on belief. But, I've argued that this claim faces serious philosophical and empirical challenges. For one thing, a central feature of beliefs is not just that accepting a proposition, but also informing inferences that we make about the world, and the actions we take; and the robust moral disposition required look an awful lot like beliefs.

The contentious nature of belief is another reason that favors moral conservationism. Suppose that (i) practical goods are secured by moral thinking and action, (ii) the nature of belief is contentious, and (iii) there aren't good reasons to discard false moral beliefs (see section 3.1). Given (i) through (iii), prudence favors moral conservationism: moral beliefs are motivational guides for us to think and act morally, and it is unclear whether robust moral dispositions would serve a comparable role, without slipping into belief.

There is good prima facie reason to suppose that we can effectively motivate someone to G (in the absence of countervailing beliefs and desires) if they believe that *they must G*. However, it is unclear if the practicing moral fictionalist can secure such practical goods without some kind of belief. And this, ceteris paribus, is good reason to favor moral conservationism.

3 | Moral Risk

3.1 | Where Moral Risk Enters

Suppose that Joanne is convinced to adopt error-theory, in part, because the objections to Mackie and Joyce fail – how should she proceed? Moral fictionalists recommend moral pretense: we should *pretend* that moral reasons hold. And this will facilitate the practical benefits of moral thinking and action without moral belief.

But there is an alternative: the serious possibility of metaethical error gives rise to moral risk – as saw earlier, there is the serious possibility of error in metaethical matters as 'the subject matter involved is the sort of thing it is all too easy for people like us to be mistaken about' (Moller 2011: 432). The idea of moral risk[33] is fairly clear: moral issues are contentious, and we risk acting in moral hazardous ways if we are mistaken in our skeptical and deflationary metaethical views.

Suppose that Joanne is worried about the moral status of buying meat. After examining vegetarian arguments, she is unconvinced. Is she morally justified in buying meat? There is reason here to think the answer is 'no'. There is an epistemic risk that Joanne is wrong about such matters. Joanne could be wrong in her reasoning in many ways like, say, the fact that our intuitions can be subject to biases like ordering and framing effects (Nado 2014). If vegetarians are right, then meat eating is seriously wrong – and this is a second-order reason for skeptics and error-theorists to adopt vegetarianism.

We should first clarify the conditions that facilitate moral risk. I lack the space to discuss moral risk *as a family of views*. And so we'll rely on the account in Moller (2011), then explain how moral risk is another reason to adopt moral conservationism.

A | AN INITIAL OBJECTION

We might worry that an appeal to moral risk wouldn't appeal to moral skeptics and error-theorists as they reject the moral reasons entirely; but moral risk at the metaethical level, appears to either *require believing in, or entail, moral realism.* Here's an analogy:

[33] There are critics of moral risk, like Weatherson (2014) and Harman (2016), but we'll bracket this off.

> John is stranded on an island in New Guinea. The locals tell him that unless he avoids eating fruit on Mondays, he'll be cursed with moongumbo, and that would be bad. After examining the evidence, John adopts moongumbo error-theory. The locals reply that even if the evidence for moongumbo is weak, John could easily be wrong about such matters — and if so, ignoring moongumbo reasons could be very bad.

But, the objection continues, John should ignore moongumbo risk – as doing otherwise would seem to require that John believes in moongumbo reasons, or entail moongumbo reasons – but he can't do that as he's an error-theorist about such reasons.

It is unclear, however, why an appeal to moral risk must either require believing in, or entail, moral realism. For instance, we could see how an atheist might pray for reasons of risk[34] like, say, an atheist in a foxhole might pray for her well-being in the hopes that it's answered. This only appears to require the non-negligible epistemic possibility that God exists; it doesn't require, or entail, theistic belief for an atheist to pray – the cost of doing so is low, theism *might be true,* and the payoff could be great if true (Kleinschmidt 2017). Likewise, the moral skeptic or error-theorist who (i) believes there is a non-negligible possibility that moral realism is true, and (ii) wants to avoid serious moral wrongdoing if moral realism is true – and if this doesn't require believing (or entail) moral realism – has a practical reason to take moral risk seriously, given the serious possibility of metaethical error (see the section on moral eliminativism).

[34] Here someone might object that prayer to God requires *belief* in God. However, Kenny (1979) argues that belief in God is not required for praying in that 'it is surely no more unreasonable than the act of a man adrift in the ocean, trapped in a cave, or stranded on a mountainside, who cries for help though he may never be heard' (129).

B | THE SERIOUSNESS CONDITION

Not every moral mistake would be serious. For example, even if Sally were wrong that telling a white lie isn't wrong, and she lied anyway, she wouldn't be doing something *that* morally wrong. However, there are cases where making a mistake would be gravely wrong. For example, if the error-theorist is mistaken in her metaethical view, but she kills her husband for the life insurance money, on most moral theories she has done something gravely wrong. The idea that moral risk should be a factor in our moral thinking depends on the seriousness of making a moral mistake. The moral risk approach doesn't say that *any* risk of a moral mistake is salient to action.

Taking moral risk as a reason to act (and believe) morally should appeal to moral skeptics and error theorists with a robust moral sense – given our robust moral sense, they might not be able to *avoid* caring about morality. Also, there are a number of issues where moral mistakes *would be serious*. If we could be seriously wrong about a moral issue – a morally weighty mistake – then that moral issue *satisfies the seriousness condition*. Only moral issues which satisfy the seriousness condition are responsive to the moral risk approach – in that sense, the moral risk approach is narrower than alternative approaches to thinking and acting morally as there are often *practical* reasons to be moral, even where there isn't any salient moral risk.

Does the seriousness condition apply to the meta-level? Refraining from thinking and acting morally would be seriously risky with regard to *practical* goods: it is plausible that one would miss out on the relevant goods (e.g. improved self-control) if they refrain from thinking and acting morally (and not that clear that it would be better *overall* to refrain from doing so). Likewise, there is a serious risk that one would be making many grave moral mistakes if they refrained from thinking and acting morally. If moral reasons didn't inform their thinking and actions, then they would be taking a serious risk of acting in a way that is seriously morally wrong (with regard to moral issues that meet the seriousness condition). It is tempting to think here that this is an argument for moral *fictionalism*: as long as we engage in moral pretense, then we can avoid acting in a manner that is morally risky – but then what we *believed* about the matter wouldn't matter. But if the most prudent way to mitigate moral risk is to preserve our moral beliefs – as beliefs are a motivational guide; but it is less clear that moral pretense would be – then we should keep our moral beliefs and eschew moral pretense.

C | THE ASYMMETRY CONDITION

Finally, moral risk emerges in cases with *asymmetrical* moral risk: where one option would be morally risky, but the alternatives would not. Suppose that Joanne thinks that it would be morally risky to euthanize a terminal infant; but also she worries that it might also be morally risky *not* to euthanize the infant. This moral issue is *symmetrically* risky: moral risk applies to every option. In contrast, there are issues like vegetarianism and abortion that are *asymmetrically* morally risky: the moral risk falls on the side of taking action (buying meat; having an abortion), but not on the side of refraining from such actions (avoiding buying meat; not terminating a pregnancy).

There is a worry lurking: there might be weightier *practical* reasons to *refrain* from acting morally than there are practical reasons *to* act morally. For example, it might be a serious moral risk to have an abortion, but there may be stronger practical reasons to have an abortion (e.g. another child would financially jeopardize one's family). At best then, mitigating moral risk is a *pro tanto* practical reason to think and act morally. The issue of countervailing goods isn't limited to the moral risk approach – the practicing moral fictionalist may have practical reasons to suspend her moral pretense (e.g. she must suspend moral pretense when lying to protect Jews in WWII).

Finally, we might wonder if the asymmetry condition applies at the meta-level. As we saw earlier, the Hobbesian presumption is good reason to think there is an asymmetry with regard to *practical* goods: if we eliminated moral reasons altogether, then we would risk losing much (if not all) of the practical goods like long-term cooperation. And there is mitigating moral risk: if we ignore moral reasons in our practical deliberations, we risk doing something very morally wrong. However, not acting on our moral beliefs might be better at mitigating moral risk. We turn to that worry next.

D | A SERIOUS OBJECTION

Perhaps the moral risk approach has similar problems as the approach in Ross (2006):

> [Suppose] that I have a degree of credence of .01 in TL [a non-nihilistic moral theory that prescribes turning the trolley to the left], but…I have a degree of credence of .99 in a nihilistic theory TN. And again suppose that I must decide between sending the trolley to the right and sending it to the left. In this case we could reason as follows. According to TL, it would be better

> for me to send the trolley to the left than to send it to the right. And so my credence in TL gives me *pro tanto* subjective reason to send the trolley to the left. The only way this could fail to be the most rational option would be if my credence in TN gave me a sufficiently strong reason to send the trolley to the right. But TN implies that there would be nothing valuable or disvaluable about either alternative. And so my credence in TN gives me no subjective reason to favor either alternative. Hence the *pro tanto* subjective reason to send the trolley to the left is unopposed … (748; original emphasis).

This line of reasoning is clear, and we can run it, using ‘moral risk’ instead of ‘subjective reason’, to arrive at a dominance argument for the moral risk approach.

The worry is that this risk argument is underspecified – call this *the underspecification objection*[35]: even if acting on skeptical or error-theoretic metaethical beliefs is morally risky sometimes, this doesn’t tell us *which moral beliefs* mitigate moral risk – and worse there are cases where ‘it might be better to believe that nothing matters’ than have moral beliefs (Peterson 2018: 600). Different moral theories offer conflicting prescriptions – on act utilitarianism, we should kill a healthy patient to save ten; but not on deontology – and remaining neutral would better mitigate moral risk than acting on the wrong moral theory, where doing so would makes things worse than not acting on such a theory.

Moral risk, however, is salient only where available moral theories agree on which actions are wrong, and their (relative) wrongness. For instance, if act utilitarianism and deontology agree on which actions are wrong, but wildly disagree on how wrong, moral

[35] For a different answer to the underspecification problem, see Tarsney (2019).

risk doesn't apply – but if they agree on which actions are wrong, and their degree of wrongness, then they are candidates for the moral risk approach. There are cases where mainstream moral theories agree. For example, many utilitarians claim that their view can ground commonsense moral prescriptions. And if their view can ground common sense prescriptions, they don't threaten utilitarianism (Hooker 2000).

A critic might worry, however, that agreement by mainstream moral theories would only mitigate moral risk if such theories are likely to be true, compared to the possible moral theories and their radically different moral prescriptions within the 'universe of logically possible evaluative judgments' (Street 2006: 122). Moral propositions like 'we should torture kids for fun,' while not prescribed by mainstream moral theories, could still be true; they have a non-zero probability. Given the moral risk approach is supposed to work, in the absence of moral knowledge, we can't assume that agreement by mainstream moral theories reflects something about the moral facts. And the evidence from the cognitive science literature implies that, once considered, we believe propositions from bizarre moral theories like, say, reverse act utilitarianism ('we should maximize pain, and minimize pleasure') to a degree. And if we have mainstream and bizarre moral beliefs, acting on our moral beliefs might not mitigate moral risk – our moral beliefs could as easily exacerbate moral risk, as mitigate it.

We can grant much of the underspecification objection, without conceding its success. It is plausible that, once considered, we hold bizarre moral beliefs like, say, reverse act utilitarianism, to a degree. And given that mainstream and bizarre moral beliefs conflict with each other, acting on our moral beliefs wouldn't mitigate moral risk. We shouldn't act on moral prescriptions from bizarre moral theories, however; we should ignore them for a couple reasons.

First, given most people's robust moral sense (see the section 'a robust moral sense'), it wouldn't facilitate long-term cooperation, say, to act on bizarre moral beliefs like 'we ought to maximize pain, and minimize pleasure' – people likely wouldn't cooperate with someone acting on bizarre moral beliefs as that would conflict with their sense of justice, goodness, and so on. And this would in turn undermine many practical gains of keeping one's mainstream moral beliefs. While we might take on various bizarre moral beliefs (once considered), and mainstream moral beliefs, we should only act on the latter beliefs – doing otherwise would threaten practical goods secured by moral thinking and action, even if a bizarre moral theory is true. (Also we can't act on mainstream and bizarre moral theories as they often offer conflicting moral prescriptions like, say, act utilitarianism and reverse act utilitarianism).

Second, the moral risk approach doesn't mitigate moral risk across the moral domain. The moral risk approach mitigates moral risk relative to mainstream moral theories – mitigating moral risk relative to mainstream moral theories, not risk in the moral domain overall, while also preserving the practical goods of thinking and acting morally like, say, long-term cooperation. By analogy, a life raft mitigates the risk of drowning, not the risk of starving. Likewise, the moral risk approach mitigates the risk that mainstream theories are right, and preserves practical goods like enhanced self-control – rather than, say, mitigating moral risk *of any kind*.

3.2 | Does Moral Risk Favor an Approach?

In addition to the doxastic troubles facing moral fictionalism, there is another reason to keep our moral beliefs: keeping our moral belief is a better approach to mitigating moral risk than pretense. Recall Joanne, an error-theorist who wants to secure practical goods like, say, improved self-control, but also mitigate moral risk – she has a prudential reason to keep her moral beliefs. If Joanne *believed* that meat eating was morally wrong, she would be motivated to eat a vegetarian diet – and help to mitigate the relevant moral risk – as beliefs are a good motivational guide.

The motivational story we told about securing *practical* goods also applies to mitigating moral risk: moral belief would motivate us to think and act morally – it is unclear that moral *pretense* would. And the best way is to think and act morally is to preserve our moral beliefs. A critic might object that there are cases where, to avoid risk, it is better to not believe anything. Here are a couple of examples:

> 1. Sally buys fire insurance for her apartment, but doesn't apparently believe that her apartment will burn – this is sensitivity to risk without belief.
>
> 2. Sam weighs two conflicting moral theories, T and T*. If he believes T, then he can't adequately respond to the risk of T* being true (and vice versa).

In reply to (1): evidence from cognitive science shows that the ‘activation of a mentally represented truth apt proposition leads immediately to believing it’ (Mandelbaum 2014: 55). So, Sally weighing the proposition (‘it is possible, but not likely, that my apartment will burn’) is enough to produce a belief (to some degree) in that proposition.

In response to (2): cases where belief in one moral theory cripples our ability to respond to risk from a conflicting moral theory are covered by the asymmetry condition – moral risk only applies where mainstream moral theories agree on wrong actions, and their degree of wrongness. It doesn’t apply to moral beliefs *writ large*, as that way leads to the underspecification objection.

4 | Conclusion

Some philosophers hold that we should forgo moral thinking and action altogether (*moral eliminativism*). This proposal runs up against the Hobbesian presumption: thinking and acting morally helps us to live better lives than we would otherwise, by securing practical goods that wouldn’t be available otherwise. Against this, some philosophers argue that we should engage in moral pretense, and discard our moral beliefs (*moral fictionalism*); or keep our (false or unjustified) moral beliefs (*moral conservationism*) – in either case, to secure the goods of thinking and action morally.

Moral fictionalism has several problems. First, there is good philosophical and empirical reasons to think discarding beliefs is harder than moral fictionalists think. Second, moral fictionalists hold that we require robust moral dispositions to engage in moral pretense;

moral pretense would be cognitively onerous without them. But such dispositions look a lot like beliefs. Finally, it would be prudent to preserve our moral beliefs: moral beliefs would be action-guiding, helping secure practical goods associated with the moral life; but it is less clear that *pretense* would be.

Finally, moral conservatism best mitigates moral risk, compared to competitors, where mainstream moral theories agree on the moral prescriptions and degree of wrongness for their violation; moral fictionalism, in contrast, is a less suitable approach to mitigating such moral risk.

REFERENCES

Alston, William P. (1988). The Deontological Conception of Epistemic Justification. *Philosophical Perspectives* 2: 257—299.

Bedke, Matthew (2014). No Coincidence? *Oxford Studies in Metaethics* 9: 102—125.

Beebe, James and Sackris, David (2016). Moral Objectivism Across the Lifespan. *Philosophical Psychology* 29 (6): 912—929.

Bloomfield, Paul (2018). Tracking Eudaimonia. Philosophy, Theory, and Practice in Biology 10 (2).

Bogardus, Tomas (2016). Only All Naturalists Should Worry About Only One Evolutionary Debunking Argument. *Ethics* 126 (3): 636-661.

Boudry, Maarten & Vlerick, Michael (2014). Natural Selection Does Care about Truth. *International Studies in the Philosophy of Science* 28 (1): 65-77.

Braddock, Matthew (2017). Debunking Arguments from Insensitivity. *International Journal for the Study of Skepticism* 7 (2): 91—113.

--- (2016). Evolutionary Debunking: Can Moral Realists Explain the Reliability of Our Moral Judgments? *Philosophical Psychology* 29 (6): 844-857.

Carruthers, Peter (1992). *Human Knowledge and Human Nature: A New Introduction to an Ancient Debate*. Oxford University Press.

Case, Spencer (2018). From Epistemic to Moral Realism. *Journal of Moral Philosophy*, pp. 1—22.

Christensen, David (2011). Disagreement, Question-Begging, and Epistemic Self-Criticism. *Philosophers' Imprint* 11.

Clarke-Doane, Justin (2012). Morality and Mathematics: The Evolutionary Challenge. *Ethics* 122 (2): 313—340.

Cline, Brendan (2018). The Tale of a Moderate Normative Skeptic. *Philosophical Studies* 175 (1): 141—161.

Cowie, Christopher (2018). Companions in Guilt Arguments. *Philosophy Compass* 13 (11): e12528.

Cuneo, Terence (2007). *The Normative Web: An Argument for Moral Realism*. Oxford University Press.

Das, Ramon (2016). Evolutionary Debunking of Morality: Epistemological or Metaphysical? *Philosophical Studies* 173 (2): 417—435.

de Cruz, Helen; Boudry, Maarten; de Smedt, Johan & Blancke, Stefaan (2011). Evolutionary Approaches to Epistemic Justification. *Dialectica* 65 (4): 517-535.

Deem, Michael J. (2018). A Flaw in the Stich-Plantinga Challenge to Evolutionary Reliabilism. *Analysis* 78 (2): 216—225.

DiSessa, Andrea A. (1982). Unlearning Aristotelian Physics: A Study of Knowledge-Based Learning. *Cognitive Science* (6): 37-75.

Dwyer, Susan; Huebner, Bryce & Hauser, Marc D. (2010). The Linguistic Analogy: Motivations, Results, and Speculations. *Topics in Cognitive Science* 2 (3): 486—510.

Elga, Adam (2007). Reflection and Disagreement. *Noûs* 41 (3): 478—502.

Enoch, D. 2010. The Epistemological Challenge to Metanarratives Realism: How Best to Understand It, and How to Cope With It. Philosophical Studies (148): 413–38.

--- (2009). How is Moral Disagreement a Problem for Realism? *Journal of Ethics* 13 (1): 15—50.

Eriksson, Björn & Olson, Jonas (2019). Moral Practice after Error Theory: Negotiationism. In Richard Joyce & Richard Garner (eds.), *The End of Morality: Taking Moral Abolitionism Seriously*. Routledge. pp. 113-130.

Fales, Evan (2009). Darwin's Doubt, Calvin's Calvary. In Michael Ruse (ed.), *Philosophy After Darwin: Classic and Contemporary Readings*. Princeton University Press. pp. 309—322.

Faraci, David (2015). A Hard Look at Moral Perception. *Philosophical Studies* 172 (8): 2055—2072.

FitzPatrick, William J. (2015). Debunking Evolutionary Debunking of Ethical Realism. *Philosophical Studies* 172 (4): 883—904.

--- (2014). Why There is No Darwinian Dilemma for Ethical Realism. In Michael Bergmann & Patrick Kain (eds.), *Challenges to Moral and Religious Belief*. Oxford University Press: p. 237-255.

Flowerree, A. K. (forthcoming). Epistemic Schmagency? In Christos Kyriacou & Robin McKenna (eds.), *Metaepistemology: Realism & Antirealism*. Palgrave Macmillan.

Frances, Bryan (2016). Worrisome Skepticism about Philosophy. *Episteme* 13 (3): 289—303.

Fraser, Benjamin James (2014). Evolutionary Debunking Arguments and the Reliability of Moral Cognition. *Philosophical Studies* 168 (2): 457-473.

Garner, Richard (2007). Abolishing Morality. *Ethical Theory and Moral Practice* 10 (5): 499—513.

Gifford, Matthew B. (2013). Skepticism and Elegance: Problems for the Abductivist Reply to Cartesian Skepticism. *Philosophical Studies* 164 (3): 685—704.

Goldman, Alvin I. (1986). *Epistemology and Cognition*. Harvard University Press.

Haidt, Jonathan and Craig Joseph (2004). Intuitive Ethics: How Innately Prepared Intuitions Generate Culturally Variable Virtues. *Daedalus*, pp. 55—66.

Harman, Elizabeth (2016). Morally Permissible Moral Mistakes. *Ethics* 126 (2): 366—393.

Harman, Gilbert (1977). *The Nature of Morality: An Introduction to Ethics*. Oxford University Press.

Heinz, Adrienne, Elizabeth Disney, David Epstein, Louise Glezen, Pamela Clark, Kenzie Preston (2010). A Focus-Group Study on Spirituality and Substance-Abuse Treatment. *Substance Use & Misuse* 45 (1-2): 134—153.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., & Gintis, H. (2004). *Foundations of Human Sociality*. Oxford University Press.

Henrich, Joe (2015). *The Secret to Our Success.* Princeton University Press.

Huemer, Michael (2016a). Serious Theories and Skeptical Theories: Why You Are Probably Not a Brain in a Vat. Philosophical Studies 173 (4): 1031-1052.

Huemer, Michael (2016). A Liberal Realist Answer to Debunking Skeptics: the Empirical Case for Realism. *Philosophical Studies* 173 (7): 1983—2010.

--- (2013). An Ontological Proof of Moral Realism. *Social Philosophy and Policy* 30 (1-2): 259—279.

Husi, Stan (2013). Why Reasons Skepticism is Not Self-Defeating. *European Journal of Philosophy* 21 (3): 424—449.

Kalderon, Mark Eli (2005). *Moral Fictionalism.* Clarendon Press.

Kelly, Thomas (2010). Peer Disagreement and Higher-Order Evidence. In Alvin I. Goldman & Dennis Whitcomb (eds.), *Social Epistemology: Essential Readings*. Oxford University Press, pp 183—217.

Kenny, Anthony (1979). *The God of Philosophers*. Oxford University Press.

Kleinschmidt, Shieva (2017). Atheistic Prayer. *Faith and Philosophy* 34 (2): 152—175.

Kyriacou, Christos (2016). Are Evolutionary Debunking Arguments Self-Debunking? Philosophia 44 (4): 1351—1366.

Jong, Jonathan & Visala, Aku (2014). Evolutionary Debunking Arguments against Theism, Reconsidered. *International Journal for Philosophy of Religion* 76 (3): 243—258.

Joyce, Richard (forthcoming). Moral Fictionalism: How to Have Your Cake and Eat it Too. In Richard Garner and Richard Joyce (eds.), *The End of Morality: Taking Moral Abolitionism Seriously.* Routledge.

--- (2008). Précis of 'The Evolution of Morality.' *Philosophy and Phenomenological Research* 77 (1): 213—218.

--- (2006). The Evolution of Morality.

--- (2001). *The Myth of Morality*. Cambridge University Press.

Kornblith, Hilary (2001). *Knowledge and its Place in Nature*. Oxford University Press.

Kyriacou, Christos (2016). Are Evolutionary Debunking Arguments Self-Debunking? *Philosophia* 44 (4): 1351—1366.

Lackey, Jennifer (2013). Disagreement and Belief Dependence: Why Numbers Matter. In David Christensen & Jennifer Lackey (eds.), *The Epistemology of Disagreement: New Essays*. Oxford University Press: pp. 243-268.

Lillehammer, Hallvard (2013). The Companions in Guilt Strategy. In Hugh LaFollette (ed.), *The International Encyclopedia of Ethics.*

Lutz, Matt (2018). What Makes Evolution a Defeater? *Erkenntnis* 83 (6): 1105—1126.

Machery, Edouard & Mallon, Ron (2010). Evolution of Morality. In John Michael Doris (ed.), *The Moral Psychology Handbook*. Oxford University Press. pp. 3—46.

Mackie, J. L. (1980). *Hume's Moral Theory*. Routledge and Kegan Paul.

Mandelbaum, Eric (2014). Thinking is Believing. *Inquiry* 57 (1): 55—96.

Mandelbaum, Eric & Quilty-Dunn, Jake (2015). Believing without Reason, or: Why Liberals Shouldn't Watch Fox News. *The Harvard Review of Philosophy* (22): 42—52.

Matheson, Jonathan (2016). Moral Caution and the Epistemology of Disagreement. *Journal of Social Philosophy* 47 (2): 120—141.

McBrayer, Justin P. (2014). The Wager Renewed: Believing in God is Good for You. *Science, Religion, and Culture* 1 (3): 130-140.

Moller, Dan (2011). Abortion and Moral Risk. *Philosophy* 86 (3): 425—443.

Moon, Andrew (2017). Debunking Morality: Lessons from the EAAN Literature. Pacific Philosophical Quarterly 98 (S1): 208—226.

Morton, Justin (2016). A New Evolutionary Debunking Argument against Moral Realism. *Journal of the American Philosophical Association* 2 (2): 233—253.

Nado, Jennifer Ellen (2014). Philosophical Expertise. Philosophy Compass 9 (9): 631—641.

Olson, Jonas (2011). Getting Real about Moral Fictionalism 1. *Oxford Studies in Metaethics* 6: 181—204.

Parfit, Derek (2006). Normativity. *Oxford Studies in Metaethics* (1): 325-80.

Pollock, John (1986). *Contemporary Theories of Knowledge.* Rowman and Littlefield.

Plantinga, Alvin (1993). *Warrant and Proper Function*. Oxford University Press.

Potochnik, Angela (2017). *Idealization and the Aims of Science*. Chicago: University of Chicago Press.

Rachels, James (1990). *Created from Animals: The Moral Implications of Darwinism.* Oxford University Press.

Reisner, Andrew (2018). Pragmatic Reasons for Belief. In Daniel Star (ed.), *The Oxford Handbook of Reasons and Normativity*. Oxford University Press, pp. 705—729.

Rowland, Richard (2016). Rescuing Companions in Guilt Arguments. *Philosophical Quarterly* 66 (262): 161—171.

--- (2013). Moral Error Theory and the Argument from Epistemic Reasons. Journal of Ethics and Social Philosophy 7 (1): 1-24.

Sayre-McCord, Geoffrey (1988). 'The Many Moral Realisms' in Geoffrey Sayre-McCord (ed.), *Essays on Moral Realism*. Cornell University Press.

Schwitzgebel, Eric & Moore, Alan T. (2015). Experimental Evidence for the Existence of an External World. *Journal of the American Philosophical Association* 1 (3): 564—582.

Schechter, Joshua (2018). Explanatory Challenges in Metaethics. In Tristram McPherson & David Plunkett (eds.), *Routledge Handbook of Metaethics*. Routledge. pp. 443—459.

Schoenfield, Miriam (2013). Permission to Believe: Why Permissivism Is True and What It Tells Us About Irrelevant Influences on Belief. *Noûs* 47 (1): 193—218.

Shafer-Landau, Russ (2012). Evolutionary Debunking, Moral Realism, and Moral Knowledge. *Journal of Ethics and Social Philosophy* 7 (1).

--- (2009). A Defence of Categorical Reasons. *Proceedings of the Aristotelian Society* 109 (1—2): 189-206.

--- (2003). *Moral Realism: A Defence*. Oxford University Press.

Singer, Peter (2005). Ethics and Intuitions. *The Journal of Ethics* 9 (3-4): 331—352.

Sinnott-Armstrong, Walter (2006). *Moral Skepticisms*. Oxford University Press.

Sobel, David & David Copp, (2001). Against Direction of Fit Accounts of Belief and Desire. *Analysis* 61 (1): 44—53.

Stich, Stephen (1993). *The Fragmentation of Reason*. MIT Press.

Street, Sharon (2009). Evolution and the Normativity of Epistemic Reasons. *Canadian Journal of Philosophy* 39 (S1): 213—248.

--- (2006). A Darwinian Dilemma for Realist Theories of Value. *Philosophical Studies* 127 (1): 109—166.

Streumer, Bart (2013). Can We Believe the Error Theory? *Journal of Philosophy* 110 (4): 194—212.

Sturgeon, Nicholas L. (1986). Harman on Moral Explanations of Natural Facts. *Southern Journal of Philosophy* 24 (S1): 69—78.

Sylvan, Kurt (2016). Epistemic Reasons I: Normativity. *Philosophy Compass* 11 (7): 364—376.

Toner, Christopher (2011). Evolution, Naturalism, and the Worthwhile: A Critique of Richard Joyce's Evolutionary Debunking of Morality. *Metaphilosophy* 42 (4): 520—546.

Turri, John (2009). The Ontology of Epistemic Reasons. *Noûs* 43 (3): 490—512.

van Inwagen, Peter (1998). Modal Epistemology. *Philosophical Studies* 92 (1): 67—84.

Vavova, Katia (2018). Irrelevant Influences. *Philosophy and Phenomenological Research*: 134-152.

--- (2014). Debunking Evolutionary Debunking. *Oxford Studies in Metaethics* (9): 76—101.

Vogel, Jonathan. (2005). Can Skepticism Be Refuted. In Steup Matthias & Sosa Ernest (eds.), *Contemporary Debates in Epistemology*. Blackwell, pp. 72—84.

Weatherson, Brian (2014). Running Risks Morally. Philosophical Studies 167 (1): 141—163.

Wedgwood, Ralph (2014). Moral Disagreement among Philosophers. In Michael Bergmann & Patrick Kain (eds.), *Challenges to Moral and Religious Belief: Disagreement and Evolution*. Oxford University Press, pp. 23—39.

Weinberg, Jonathan M.; Gonnerman, Chad; Buckner, Cameron & Alexander, Joshua (2010). Are Philosophers Expert Intuiters? *Philosophical Psychology* 23 (3): 331—355.

Wielenberg, Erik J. (2016). Ethics and Evolutionary Theory. *Analysis* 76 (4): 502-515.

--- (2009). In Defense of Non-Natural, Non-Theistic Moral Realism. *Faith and Philosophy* 26 (1): 23—41.