MONASH University

Moral Naturalism as a Response to Evolutionary Debunking Arguments

Matthew Robert Ringenbergs

BA(Hon)/BEc

A thesis submitted for the degree of Master of Arts at

Monash University in 2019

Department of Philosophy

Copyright notice

Abstract

Evolutionary Debunking Arguments (EDAs) attempt to undermine our justification in a set of beliefs from the fact that the beliefs in question are the result of evolutionary processes. The EDA against morality attempts to show that our moral beliefs are likely the result of an evolutionary process that would be insensitive to moral truth (that is the process would be "off-track"), and thus it would be a huge coincidence for our ordinary moral beliefs to be true, undermining our justification for accepting those moral beliefs. However, justification can be rescued if a compelling moral naturalist account can be provided; if we have reason to believe that the moral facts are identical to, or grounded in, the natural facts that play an explanatory role in our moral belief-formation process, then we have reason to believe that our moral beliefs are sensitive to the truth after all.

Two proponents of EDAs against morality, Joyce (2006) and Street (2006, 2008), provide arguments that attempt to show that such a compelling naturalist theory is not extant and likely never will be. Joyce makes the claim that moral values *a*re inescapably authoritative, that is they have 'practical clout', and that given the nature of moral values as such, no moral naturalist can provide the right naturalist account, as no moral naturalist theory can account for practical clout. For Joyce (2006), revisionary approaches that attempt to abandon practical clout fail to be compelling because they fail to count as theories about *morality*. Meanwhile, Street (2006, 2008) argues that the value naturalist is begging the question.

This thesis argues that moral naturalist theories that meet certain reasonable constraints can avoid the force of these challenges, and thus meet the epistemological challenge posed by the EDA. By drawing on Joyce's (2005) argument for revisionary moral fictionalism, I will also introduce a variation of certain types of revisionary moral naturalist theories that may help in this endeavour, which I will call fictional-internalist externalism, or FI-externalism for short. Essentially, the FI-externalist, in their most critical contexts, adopts a moral naturalist theory that abandons practical clout, but while in ordinary contexts, they go on 'make-believing' in practical clout in order to gain certain regulative benefits. I argue that the FI-externalist approach holds several advantages over both standard moral naturalist and moral fictionalist approaches. In particular FI-externalism overcomes a major objection to the ability of standard moral naturalist approaches to meet the epistemological challenge of the EDA.

Declaration

This thesis is an original work of my research and contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Print Name: Matthew Ringenbergs

Date: 31/05/2019

Acknowledgements

I would like to express a great thank you to my supervisor Professor Toby Handfield, for his extensive and detailed feedback and guidance throughout the thesis writing process. Thank you as well to my associate supervisor Dr. John Thrasher for his comments and constructive criticism. I would also like to thank my friend, Shang Long Yeo, for helpful discussion and direction to useful articles. Finally, thank you to my parents for all their encouragement and support. This research was supported by an Australian Government Research Training Program (RTP) Scholarship.

Table of Contents:

Introduction

Evolutionary Debunking Arguments (EDAs) attempt to undermine our justification in believing a particular belief or set of beliefs on the grounds that the belief or set of beliefs in question are the result of evolutionary processes. In particular, EDAs have often been brought to bear on the moral domain. The EDA against morality attempts to show that our moral beliefs are likely the result of an evolutionary process that would be insensitive to moral truth, that is it attempts to show that these evolutionary processes are likely "off-track" and likely to produce beliefs that do not correlate to the truth. Thus, the EDA argues that it would be a huge coincidence for our ordinary moral beliefs to be true and we therefore have no justification for accepting those moral beliefs. Two of the more well-known EDA's for morality have been proposed by Joyce (2006) and Street (2006).

Just because the EDA may reveal our moral beliefs to be unjustified, does not mean justification cannot be rescued[1]. The EDA properly construed specifically targets the epistemic conservative, who argues that their most firmly held beliefs should be considered justified until shown otherwise i.e. "innocent until proven guilty" (Joyce, 2016a, 2016d). Moral success theorists, who argue that at least some of our moral beliefs are justified, must therefore either a) establish that the EDA is wrong and we have no reason for doubt, or (b) that it is right and we do have such a reason, but that we also have the resources to dismiss it. For example, the latter approach might be achieved if a *compelling* moral naturalist account can be given. If we have reason to believe that the moral facts are identical to, or grounded in, the natural facts that play an explanatory role in our moral belief-formation process, then we have reason to believe that our moral beliefs are sensitive to the truth after all. Of course providing a moral naturalist account that is compelling is a difficult challenge, and both Joyce (2006, 2016d) and Street (2006, 2008) provide arguments that attempt to show that such a compelling naturalist theory is not extant and likely never will be. Joyce makes the claim that moral values are *inescapably authoritative,* that is they have 'practical clout', and that given the nature of moral values as such, no moral naturalist can provide the right naturalist account, as no moral naturalist theory can account for practical clout. While Street (2006, 2008) argues that the value naturalist is either begging the question or is subject to the same EDA 'one level up'. The aim of this thesis is to show how otherwise compelling moral naturalist theories can avoid the force of these challenges, and thus meet the epistemological challenge posed by the EDA.

[1] Or shown to be justified all along depending on your view of justification.

To be clear, the aim of this thesis is not to argue for the plausibility of any particular moral naturalist theory, but to make it clear that the force of the arguments that Joyce and Street level against moral naturalism can be largely avoided by certain types of moral naturalist theories.

This thesis will have three chapters. In Chapter 1, I will provide an overview of the relevant literature. I will first introduce and explain the EDA's of Joyce (2006) and Street (2006), describing several key similarities. I will then provide an overview of the relevant differences in moral naturalist theories. I will finish by describing the objections that Joyce (2006) and Street (2006, 2008) put to moral naturalism.

In chapter 2, I will discuss Joyce's (2006) argument against moral naturalism in more detail, as well as another argument he makes, his argument for moral fictionalism as a response to a moral error theory, in order to show how we can use resources from this latter argument to actually undermine the former argument. Throughout this chapter I will explain what Joyce sees the value/function of morality as being, explain his argument that the moral fictionalist can achieve such benefits without believing in morality as being literally true, and show how (reasons) externalist moral naturalism with some modifications can also achieve these same benefits, thus undermining his argument that morality needs to be reasons internalist to fulfil its function. To this end I introduce a version of (reasons) externalist moral naturalism called *fictional-internalist externalism* or *FI-externalism* for short. I will also give some reasons as to why we might prefer an f-internalist theory over accepting a moral error theory and adopting moral fictionalism or abolitionism.

In Chapter 3 I will discuss Street's evaluation of the arguments of the value naturalist and her two main objections to those arguments, the "one level up" objection and the "trivially question-begging" objection. I will then explore whether moral naturalism can defeat these objections concluding that while such a theory may be able to vindicate (at least partially) moral realism, it would be unable to vindicate, on its own, evaluative realism as a whole, the theory that the evaluative truth is independent of any of our evaluative attitudes. This may not be such a great cost though. As will be seen in Chapter 2, by adopting f-internalism, we may be able to limit the practical cost of losing the inescapable authoritativeness of, and desire-independent reasons for, acting morally.

A key concern of both Joyce and Street is that by adopting moral naturalism one may appear to be losing something. For Street it is the normativity of morality. For Joyce, it is practical clout and

thus supposedly the benefits of morality. Hence, an externalist moral theory that can resist their respective EDAs may seem like a non-starter. But I hope to show, that although we may not be able to vindicate our entire original concept of morality, we can vindicate parts of it, and through adoption of an FI-externalist version of an externalist theory, the practical effects of such a loss is likely quite small. Therefore, if we have other good reasons for adopting a particular externalist naturalist theory we are better off revising our moral beliefs accordingly than suspending them completely as Joyce maintains. Hence, rather than completely undermining our justification in our moral beliefs, the EDA instead merely poses a challenge, one that can be met through revising our moral beliefs in accordance with an externalist, perhaps even f-internalist, moral naturalist theory. Moral naturalism can therefore partially vindicate morality in response to an Evolutionary Debunking Argument.

Chapter 1:

Literature Review

This chapter explores some of the literature surrounding Evolutionary Debunking Arguments against morality, and the possibility of moral naturalism as a response. The chapter will begin by describing the use of Evolutionary Debunking Arguments (EDAs) in general before moving on to the specific EDAs of Joyce (2006) and Street (2006). I will compare and contrast the two, describing key similarities, also illustrating how moral naturalist accounts may be able to avoid the debunking force of these arguments in similar ways. Finally, the chapter will discuss some of the arguments that Joyce and Street make against the possibility of a compelling moral naturalist account that can avoid the EDA. In Chapters 2 & 3, I will go on to show that certain types of moral naturalist theories can largely avoid the force of these negative arguments.

1.1. Evolutionary Debunking Arguments

Evolutionary Debunking Arguments (EDAs) have been conceived as a new version of a much older type of argument, the Debunking Argument (Kahane, 2011). Debunking arguments are arguments that use the causal origin of a belief or set of beliefs to show that they are unjustified[2]. These arguments rely on two main premises. The first, the "causal premise", is that the belief is formed by some mechanism, and the second, the "epistemic premise", is that the formation of beliefs by this mechanism is not sensitive to the truth, that is, the process does not track the truth and thus is unreliable (in Kahane's words, it is an "off-track process"). These two premises together entail that the belief in question is unjustified. This is not to say that the belief is false as the argument does not give reason to *disbelieve*, just no reason to *believe* (Joyce, 2006; Kahane, 2011).

For the EDA, the mechanism involved is natural selection. The idea is that in some cases, natural selection may have contributed to the formation of beliefs, not on the basis of their veracity, but on how much they improved fitness. In many cases, these two criteria will align; believing what is true often leads to greater fitness, whereas believing incorrect things will often lead to lesser fitness. However, there are some cases where the act of holding a belief may be fitness enhancing even

[2] Whether the EDA shows that such beliefs are no longer justified or were never justified in the first place is unimportant for this thesis. The EDA functions the same regardless of the theory of justification.

without the belief being true. In such cases then, the EDA may apply, since the mechanism that formed the belief, natural selection, is not sensitive to the truth, and thus does not track the truth. There have been a number of attempts to show that at least some moral beliefs and intuitions fit this profile, and, consequently, that holding such beliefs or intuitions is unjustified. For example, Greene (2008) and Singer (2005) have both attempted to show that our deontic intuitions are the result of off-track evolutionary influence and, therefore, reliance on such intuitions is unjustified. Other EDAs are broader, targeting all moral beliefs and intuitions, or even all our realist beliefs about value (thereby targeting realist beliefs about morality too). This thesis will focus on the latter, 'global', Evolutionary Debunking Arguments.

There is some evidence for the hypothesis known as moral nativism, the theory that the disposition to make moral judgments is innate to humans and largely the result of natural selection, being fitness enhancing through, primarily, the ability to promote cooperative behaviour (Joyce, 2006; Sober & Wilson, 1998). Since the disposition to make moral judgments is likely to be fitness enhancing regardless of whether or not such judgments are true (even moral nihilists can endorse moral nativism), there is room to use EDAs to show that at least some of our moral beliefs are unjustified. Two of the more well-known and successful EDA's for morality have been proposed by Joyce (2006) and Street (2006)[3]. In the interest of space and going into depth, this thesis I will focus on these two formulations of the EDA, but other formulations include Bedke (2009), Kitcher (2011) and Ruse (1986, 2006). Many of the arguments in this thesis will apply to these other EDAs.

There has been some discussion of how exactly EDAs differ from other kinds of sceptical arguments, if they indeed do. The answer to this question is important, because if they do not differ from other, more general, sceptical arguments, then we may be able to simply dismiss them in the same ways we attempt to dismiss arguments about brains in vats or evil demons etc. However, Vavova (2015) provides a case as to why EDAs should be distinguished from and receive different responses than other sceptical arguments. In order to argue her case, she provides a good summary of Street's (2006) argument for the evolutionary debunking of morality. It goes as follows:

1. Realism. Moral truths are attitude-independent.
2. Influence. Evolutionary forces have influenced our moral beliefs.

[3] It should be noted that Street's (2006) EDA is supposed to target evaluative realism as a whole, the view that there are at least some evaluative facts (facts about values or reasons for action) that hold independently of any and all of our evaluative attitudes. Since moral facts are generally considered to be a kind of evaluative facts, Street's EDA thereby also appears to target moral realism.

3. Off-track. Evolutionary forces aim at fitness, not attitude-independent moral truths.
4. Gap. The fitness enhancing beliefs and the moral truths come apart.
5. Mistaken. We have good reason to think that our moral beliefs are mistaken. (Vavova, 2015; p. 108)

Vavova does this because she wants to distinguish it from another argument that is often drawn from Street's (2006) paper. This latter evolutionary debunking argument is as follows:

1. There are many possible coherent normative belief systems.
2. Only one of these is right.
3. The odds are phenomenally low that mine is the right one.
4. I have no non-question-begging evidence that mine is the right one.
5. If the odds are low that I'm right and if I have no non-question-begging evidence that I'm right, I cannot conclude that I'm right.
6. I cannot conclude that my normative belief system is the right one. (Vavova, 2015; p. 107)

For Vavova (2015), this latter argument, while seemingly persuasive, is not any more effective in specifically debunking moral realism than other generally sceptical arguments. Vavova's aim is to distinguish EDAs from generally sceptical ones. She argues that we often simply dismiss the latter, largely because the possibility of error raised by the sceptic is merely possible, not empirical (Vavova, 2015; p. 105). However, Vavova argues that a properly construed EDA cannot be so easily dismissed, at least not in the same ways. To properly construct a hypothetical EDA such that this is the case, she suggests it must have three features. First, the argument must be empirical; it must do more than raise the possibility of error, it needs to make such a possibility probable by introducing actual evidence of error. This is much like how an optometrist noting that the tests they have run indicate that you are colour-blind is different to a sceptical argument that claims that because there is a possibility that all of your colour judgements are in error then you are unjustified in believing your judgments are correct. The evidence of error makes the optometrist's arguments harder to dismiss, unlike with a sceptic. Second, the argument needs to be targeted, that is it should only threaten moral beliefs, not extend to other areas such as mathematics. Lastly, it should be epistemological, that is it should not conclude that there are no moral truths, but that we cannot (currently) know them.

Vavova (2015) believes that the latter interpretation of Street's (2006) argument fails to meet these desiderata; it references no evidence to suggest that the possibility of error is more than just a possibility, and it can be applied just as easily to all sorts of beliefs (for example, Clarke-Doane (2012) shows how such an argument can be extended to apply to realist mathematical beliefs). In these regards, it is no different than other sceptical arguments. Bedke's (2009) EDA, an argument from cosmic coincidence also seems to fit this profile. The former argument, however, has all three features according to Vavova. It is empirical, noted by Vavova to be truly evolutionary in the sense that it provides evidence from evolutionary theory that our moral beliefs are likely off-track. It is targeted since accuracy is not adaptive for our moral beliefs but may be for some of our empirical beliefs, thus the EDA cannot be extended into areas we do not want debunked. Finally, the argument is epistemological; it aims to show you that your confidence in your moral beliefs is unwarranted, that by your own lights you are probably mistaken. This means that for an EDA to be successful and significantly different to general sceptical arguments, it needs to be more like the former argument than the latter. Both the EDA's of Joyce (2006) and Street (2006) can be read as fitting the former pattern rather than the latter.

The evolutionary nature of the EDAs, such as those of Street (2006) and Joyce (2006), therefore plays a key role in the force of their arguments. The moral nativist premise ensures that, at the very least, these arguments are empirical and epistemic, and thus can be distinguished from other, more general sceptical arguments. The properly construed EDA requires a different response than the sceptic; it is not enough to just provide a moral epistemology, how we *could* have come to know the moral facts. Instead, the response must either (a) establish that the EDA is wrong and we have no reason for doubt, or (b) that it is right and we do have such a reason, but that we also have the resources to dismiss it. The EDA properly construed is therefore a burden shifting argument, a challenge to the moral success theorist. It specifically targets the epistemic conservative, who argues that their most firmly held beliefs should be considered justified until shown otherwise i.e. "innocent until proven guilty" (Joyce, 2016a, 2016d). General sceptical arguments which only show the possibility of error do not touch the epistemic conservative. Meanwhile the EDA, properly construed, aims to show actual evidence of error, proving the firmly held beliefs "guilty", or at least not "innocent". The burden is then on the moral success theorist to show either that the EDA is wrong, or to give an account that provides evidence that their moral beliefs are justified.

There have been a number of attempts to show that we have exactly these resources for overcoming the debunking argument at our disposal (for example, Copp, 2008; FitzPatrick, 2014;

Sterelny and Fraser, 2016; Shafer-Landau, 2012). A commonly proposed solution draws on moral naturalism, the realist view that not only are there at least some moral facts that hold independently of any and all of our evaluative attitudes, but that all of these moral facts are either identical to or entirely grounded in natural facts. Such views attempt to avoid the debunking force of the EDA by linking the evolutionarily influenced attitudes and the independent evaluative/moral truth by a third factor, in this case by some set of natural facts. Berker (2014) gives the definition of a third factor account as follows:

> *A third-factor account:* Evolutionary forces have tended to make our [evaluative] judgments track the attitude-independent [evaluative] truth because, for each [evaluative] judgment influenced by evolution in this way, there is some third factor, F, such that
>
> (i) F tends to causally (help) make it the case that (proto) judging in that way promotes reproductive success (when in our ancestors' environment), and
>
> (ii) F tends to metaphysically (help) make it the case that the content of that judgment is true. (Berker, 2014; p. 15)

Instead of relying on a causation relation to explain why a given judgment tracks a given fact, third-factor accounts, and thus moral naturalist theories, instead make use of the grounding relation (or its converse, the in-virtue-of relation) (Berker, 2014). Essentially, such theories posit, for every normative judgment that *p,* there is "some non-normative third fact on which [the] judgment that *p* causally depends and on which the fact that *p* metaphysically depends" (Berker, 2014; p. 16). Taking a third-factor approach in this way allows one to explain how our evaluative/moral judgments can track the evaluative/moral facts without assuming the evaluative facts have causal powers (Berker, 2014). However, providing such a third-factor account, providing an explanation of what exactly grounds the evaluative/moral facts and how, is no easy task. As discussed earlier, the EDA puts the burden of proof is on the success theorist to show how our beliefs track the truth. Therefore, showing that a third-factor account is possible will not do, such an account has to actually be provided, and it needs to be an account we can believe in, one that is *compelling* (Joyce, 2006, 2016d). Both Joyce (2006, 2016d) and Street (2006, 2008) make arguments doing so is unlikely, if not impossible.

A moral naturalist view, properly construed, therefore may be able to resist the EDA. If such a view is true, if the moral facts supervene on the natural facts, and an account of how this is so can be given, then it can be shown how we evolved to track the moral/evaluative facts, because it could

be shown how we evolved to track the natural facts that ground them. So if we have a moral naturalist theory that explains how this is possible, then the 'epistemic premise', that the evolutionary forces that produced our moral beliefs are off-track, would fall through. Such is the importance of moral naturalism that Das (2016) argues that this is where the majority of the debunking force of EDAs against moral realism lies; without a metaphysical assumption that our moral beliefs are not identical or grounded in natural facts, the debunking force of the EDA is quite modest. Both Joyce (2006) and Street (2006) therefore spend some time grappling with the question of the viability of moral and/or value naturalism in their works, putting forward arguments that either claim moral naturalism cannot account for what makes a moral theory *moral,* or argue that value naturalism puts off the EDA to a higher level. There are many versions of moral naturalism though, some which may have more potential than others. This thesis aims to show that some types of moral naturalism have the capacity to avoid these additional arguments put forward by Joyce and Street. But first I want to discuss in some more detail the individual EDAs of Joyce (2006) and Street (2006).

1.2. Joyce's EDA

Richard Joyce's EDA originates in his book *The Evolution of Morality* (2006), and has been developed and discussed further in Joyce (2016a, 2016d, 2016e). Although some of the details and emphasis has changed over the years, the core of the argument has remained the same. In the book, Joyce (2006) argues for a moral nativist premise, that humans have an innate tendency to think of certain actions and omissions as morally required, as categorical imperatives, and that our possession of this trait, due to its likely fitness enhancing properties, is a result of natural evolutionary processes (the causal premise of the debunking argument). He then goes on to argue that in this case, the evolutionary processes that resulted in our tendency to form moral judgements do not appear to be sensitive to the truth (the epistemic premise). This gives us reason to suspect the reliability of our belief-formation processes for the moral domain, leading to an epistemological challenge, one that argues that unless a *compelling* and *plausible* moral naturalist theory can be provided, we have good reason to believe our moral judgments are unjustified, that we should neither believe nor disbelieve them, and that our ordinary moral judgments are systematically in error.

It should be noted that Joyce (2006, 2016a, 2016d) intends only a modest conclusion for this EDA, which attempts only to show that our moral beliefs are unjustified. In addition, it allows for justification to be rescued if a compelling moral naturalist account can be provided. To argue for a stronger conclusion, for example that all moral beliefs are false, would take extra steps involving some controversial metaethical commitments (Joyce, 2016e). It should also be noted that Joyce (2006) has argued that it would be a mistake to assume his EDA only targets morally realist beliefs; for subjectivist and constructivist views can be targeted as well (Joyce, 2016a, 2016d). So while Joyce's debunking argument may have a modest epistemological conclusion, it is quite broad, targeting the entire moral domain.

An important aspect of the EDA is that the moral nativist hypothesis does not presume that the human faculty of moral judgment served reproductive fitness via the production of *true* judgments. Instead, most moral nativist theories suggest that morality evolved due to its ability to promote cooperation, enhance social cohesion and regulate behaviour (Joyce, 2006; Joyce, 2016d; Kitcher 2011; Sterelny and Fraser, 2016), generally through such mechanism such as kin selection, mutualism, reciprocity and, more controversially, group selection (Alexander, 1987; Joyce, 2006; Sober and Wilson, 1998). For example, holding a tendency to think of cooperative behaviours as morally required regardless of one's desires is theorised to enhance motivation to cooperate with others, more so than thinking purely in terms of self-interest (Joyce, 2016, 2016d). I will discuss some of these adaptive benefits of morality in Chapter 2. The point here is that in none of these moral nativist hypotheses does the truth or falsity of those moral beliefs figure; the truth or falsity of a set of held moral beliefs appears to have no impact on whether such beliefs promote fitness in the ways described above. In fact, both the moral realist and the moral error theorist could be convinced of the plausibility of the nativist hypothesis without contradiction (Joyce, 2016d). According to this evolutionary hypothesis then, the process by which we came to form our moral beliefs, that is to say natural selection, is not truth-tracking, it is not *sensitive* to the truth.

Joyce (2016d), drawing on Gilbert Harman (1986), argues that the truth-sensitivity of a belief formation process should be understood in terms of whether the truth plays an *explanatory role* in the belief-formation process[4]. Harman argued that "what's needed is some account of *how* the

[4] Since his 2006 book, Joyce has changed position in regards to how to understand truth-sensitivity, particularly in regards to the truth-sensitivity of the process by which we came to hold moral beliefs. Joyce (2006) argues that since, according to the moral nativist hypothesis, the fitness enhancing properties of our moral beliefs does not depend on their truth or falsity, then even if there were no moral facts, we would still form moral beliefs that we perceived as categorical imperatives, and such moral beliefs would likely be similar to those we hold now. On this sort of view, for a belief-formation process to be sensitive to the truth, and thus

actual wrongness of [an action] could help explain [someone's] disapproval of it. And we have to be able to believe in this account. We cannot just make something up…" (p. 63). More specifically, Joyce argues that:

> Whether the faculty tracks the truth depends on whether the judgments covary with (or are explained by) those fact(s) that they represent—in this case, X's being P. And whether the faculty has the function of tracking that truth depends on whether success at truth-tracking explains the emergence and persistence (and thus the very existence) of the faculty (Joyce, 2016d; p. 150).

In regards to the moral domain, the moral nativist hypothesis suggests that the emergence and persistence of the relevant faculty can be wholly explained by non-truth-tracking functions, in particular the faculty's ability to promote cooperation in the environment of our ancestors (Joyce, 2016d). According to these nativist hypotheses, in order to explain how our moral judgments came to be in the ancestral environment, it is not necessary to refer to their being true; instead we can show how they proved evolutionarily advantageous. This can be contrasted with perceptual beliefs about mid-sized objects in the environment. In order to explain how we can to form such beliefs we need to refer to their truth. Forming beliefs about tigers, trees, cliffs, etc. in our environment is only advantageous if there really are those things in the environment. Thus, the truth of mid-sized objects in the environment plays an *explanatory role* in how beliefs about such objects in the environment came to be, while the facts of the matter do not appear to have played an explanatory role in how our moral beliefs came to be, considering the other scientific hypotheses and evidence we have available. Therefore, the process by which we came to form our moral beliefs, that is to say natural selection, appears to be insensitive to the truth in this case, and thus unreliable. Since finding out that a process by which you came to a belief is unreliable should undermine the justification in that belief, the EDA therefore undermines our justification for believing in moral facts.

This EDA can be seen to be a version of the more effective EDA approach outlined by Vavova (2015). It is epistemic, providing evidence from evolutionary theory that our moral beliefs are off-track, it is targeted, debunking only our moral beliefs and leaving other domains of knowledge

---

truth-tracking, then for any proposition *p,* "(i) if *p*, then S believes *p*, and (ii) if *not-p*, then S does not believe *p*" (Joyce, 2016(d); p. 147). In later works (for example Joyce (2016d)), Joyce argues that this interpretation is problematic, because it has difficulties with accounting for beliefs that concern necessary truths and necessary falsehoods, which could cause the counterpossible conditionals in the above to turn out vacuously true. As a result of this, Joyce moves away from an explanation of truth-sensitivity and truth-tracking in terms of counterfactual covariance and puts a heightened emphasis on considerations of the *explanatory role* of the truth in the belief-formation process.

intact, and it is epistemological, it aims to show only that such beliefs are unjustified. The EDA therefore cannot be dealt with in the same ways with which we deal with more general sceptical arguments. As mentioned, the conclusion is quite modest, specifically it is a burden shifting argument, placing the burden on the success theorist to show that either the EDA is unsound, or that it is succeeds and yet our moral belief formation processes track the truth anyway. If neither of these approaches can be achieved then we may be forced to accept a sceptical conclusion, that all our moral judgements are unjustified, and that the moral truth cannot be known. Since Joyce (2006) argues that ordinary moral discourse is cognitive i.e. moral sentences express moral propositions, and the acceptance of a moral sentence is the belief in the moral proposition expressed, the EDA, if successful, would suggest that such discourse is systematically in error. Unlike moral error theorists such as Mackie (1977), where the error is in believing a false proposition (because there are no true moral propositions), the error here is in believing an epistemically *unjustified* proposition (because the EDA shows that all of our moral beliefs are unjustified) (Kalderon, 2005)[5].

The modest epistemological nature of Joyce's (2006) EDA needs to be highlighted. Not only does the EDA not seek to establish that our moral beliefs are false, instead seeking to establish the weaker conclusion that our moral beliefs are unjustified, it does not seek to establish that this lack of justification is permanent. It is entirely possible that we could re-establish the justification of our moral beliefs. The target of the EDA is the epistemic conservative, who argues that we should give firmly held beliefs the benefit of the doubt, that we can assume that our moral beliefs issue from a process in which the moral facts play an explanatory role until shown otherwise i.e. 'innocent until proven guilty'. Moral nativism however provides empirical evidence that supports a genealogical hypothesis which a (nihilistic) moral error theorist can wholly endorse, threatening this assumption. This is not to say that the EDA supports moral nihilism, but it shifts the burden of proof onto the

[5] While Joyce (2006) originally followed Kalderon (2005) in labelling this position a moral error theory, he has since recanted this position (Joyce, 2016e), instead arguing that moral nihilism is necessary for moral error theory, and that the epistemological conclusion of his EDA could conceivably be compatible with moral realism. However, while it is true that a moral judgment can be true while being unjustified, Joyce (2006) has also argued that someone cannot both accept the epistemological conclusion of the EDA and still remain a moral realist without violating important epistemic principles (p. 162). Nevertheless, I do think it may be useful to separate out the disjunction in Kalderon's (2005) formulation of moral error theory. To that end we can draw a distinction between what I'll call moral agnosticism (the view that our moral judgments are unjustified), and thus we should neither believe nor disbelieve them) and moral nihilism (the view that our moral judgments are false, and thus we should disbelieve them). We can, therefore, further draw a distinction between agnostic moral error theory (the view that our moral discourse is systematically in error because it is unjustified) and nihilistic moral error theory (the view that our moral discourse is systematically in error because it is false). Joyce's (2016e) newer position labels nihilistic moral error theory as simply moral error theory, but I do not think it matters which is used as long as it is clear what is meant. Drawing a distinction like this is important because, as Joyce (2016a) notes, many commenters confuse the epistemological, agnostic error theoretic conclusion of his EDA with a nihilistic error theoretic conclusion, which would require a number of extra steps and metaethical assumptions not included in Joyce's original argument.

success theorist. The success theorist either needs to show that moral nativism is incorrect, or that the moral error theorist cannot endorse the genealogical hypothesis.

In the previous section we discussed third-factor accounts, which attempt to avoid the debunking force of the EDA by linking the evolutionarily influenced attitudes and the independent evaluative/moral truth by a third factor. The success theorist can attempt this kind of approach. For example moral naturalist approaches will attempt to ground or identify the moral facts with some set of natural facts that play an explanatory role in the belief-formation process. In this case, the moral facts would then play the necessary explanatory role in our belief-formation processes *via* the natural facts that play such an explanatory role. But such a third-factor account cannot merely be presented as a possibility to be effective, the account has to be fleshed out; we need an explanation of exactly what natural facts the moral facts are grounded in or identical to and how they play the required role in the belief-formation process, and we need good reason to accept the account, for example empirical evidence. The account also has to be believable; whatever property is picked out by the account must satisfy our desiderata for a *moral* property. I will call the former condition the *Good-Reason* constraint, and the latter, the *Desiderata* constraint. These two constraints must be satisfied for any moral naturalist theory to be considered compelling enough to be accepted as a satisfactory response to the EDA. However, while they are necessary conditions, they may not be sufficient, it is possible that other conditions must be satisfied.

Joyce (2006) does consider the possibility of third-factor accounts, by way of considering moral naturalism as a response to the EDA, invoking Harman's (1977) Challenge:

> …if there is no reductive account available explaining how moral facts relate to naturalistic facts, then moral claims cannot be tested, moral theories cannot be confirmed or disconfirmed, and we have no evidence for the existence of moral facts (Joyce, 2006, pp. 184-185)

The point Harman (1977) is trying to make, argues Joyce (2006), is that it is possible to give a complete explanation of moral judgments in which the truth or falsity of those judgments is irrelevant. Therefore, unless moral concepts can be reduced to natural, physical properties of the world, then moral concepts can be safely removed from our ontology without losing anything of substance. Thus, Joyce argues that while it may be possible to use Occam's razor to remove 'non-natural' and 'supernatural moral facts' from our ontology, it would be too hasty to argue that

Occam's razor allows us to remove 'moral facts' entirely; it is possible that these moral facts could be reduced to the non-moral or natural facts invoked by the causal explanation utilised in the debunking argument. This is in much the same way that a cat may be reduced to its physical/biological properties, but cats may be included in our ontology without being superfluous. However, a moral naturalist account must not merely be possible, it must be *compelling*; a concrete theory must be given as to how the moral fits into the natural, or we must have some reason to think such a theory is forthcoming. The fact that it is conceivable that ghostly properties could be reduced to natural ones gives us no reason to believe in ghosts. So too is it the case with morality; it is not simply enough that it is conceivable that the moral facts are grounded in natural facts, we need to have a concrete theory about how this is so, or at least we need to have evidence that one is forthcoming (Joyce, 2006, p. 189-190). It must be shown that there is *good* reason to think that the (nihilistic) moral error theorist cannot endorse the genealogical hypothesis. So Harman's Challenge is to provide a compelling account of how moral concepts can be reduced to natural ones, or to show that one is forthcoming. Therefore, for the EDA to be successful, it must be shown that Harman's Challenge has been met.

Joyce (2006, 2016d) is very clear that he does not think that any such compelling account is currently available. As a necessary condition (but not a sufficient one) for a moral naturalist theory to count as a compelling theory, Joyce (2006, pp. 190-191) argues that it must satisfy certain requirements or constraints regarding what moral phenomena must be like. The Desiderata constraint must be satisfied. Moral properties cannot just be anything, any natural property. It cannot be for instance that morally good actions are simply whatever improves the possibility of us passing on our ancestors' genes, as that would not satisfy what we consider morality to *be*. Joyce (2006, 2016d) thinks no extant theory can satisfy our moral desiderata satisfactorily, and furthermore, that it is unlikely any ever will.

So what are these moral desiderata, the platitudes that enough of which any system of morality must account for to be considered a *moral* theory? Joyce (2006) argues that it is hard to say, perhaps even impossible. Having said that, for Joyce, the key platitude, the one which moral naturalist theories *absolutely* must account for in order to be theories about *morality*, is the perceived *inescapable authority of moral judgments*, also called the 'practical clout' of morality. Essentially, Joyce argues that people see moral propositions and judgments as being *inescapable*, that is they apply universally to everyone without exception, and *authoritative*, they provide reasons on their own to comply with the moral proposition in question, independent of any of our desires (a

thesis known as internalism). Joyce suggests that it is hard to see how naturalistic facts could possibly account for the practical clout of morality, and thus argues that providing a compelling moral naturalist case that satisfies Harman's Challenge would be unlikely.

I will discuss further the arguments Joyce (2006) makes in arguing for the necessity of practical clout and the inability for moral naturalism to account for it in Section 1.5. In the next section though, I will discuss the EDA constructed by Street (2006), and compare it to Joyce's.

1.3. Street's EDA

In comparison to Joyce (2006), the target of Street's (2006) EDA is far broader, extending beyond the moral realm to all evaluative concepts, even to the extent of epistemic principles (Street, 2009). In particular, her targets are contemporary realist theories of value that claim to be compatible with natural science. She argues that Darwinian considerations pose a dilemma for these theories. Essentially, while evolutionary forces have had a tremendous influence on the content of human evaluative attitudes[6], realist theories of value posit that there are at least some evaluative facts or truths[7] that hold independently of all our evaluative attitudes. Therefore, the challenge for such theories is to explain the relation, if any, between these evolutionarily influenced attitudes and these independent evaluative truths. Street argues that realism can give no satisfactory account for this relation. Realists have two options they can take; they may either assert the existence of such a relation or deny it.

If realists deny a relation between evolutionarily influenced attitudes and independent evaluative truths, then the forces of natural selection must be viewed as a purely distorting influence on our evaluative judgments. Street (2006) argues that it would be an implausibly large coincidence that the majority of our evaluative judgments (that happened to be shaped by the distorting

[6] From Street (2006): "*Evaluative attitudes* I understand to include states such as desires, attitudes of approval and disapproval, unreflective evaluative tendencies such as the tendency to experience X as counting in favor of or demanding Y, and consciously or unconsciously held evaluative judgements, such as judgements about what is a reason for what, about what one should or ought to do, about what is good, valuable, or worthwhile, about what is morally right or wrong, and so on" (p. 110).

[7] From Street (2006): "*Evaluative facts* or *truths* I understand as facts or truths of the form that X is a normative reason to Y, that one should or ought to X, that X is good, valuable, or worthwhile, that X is morally right or wrong, and so on" (p. 110).
N.B: 'Evaluative' and 'normative' are often used interchangeably in the literature; 'evaluative' tends to put the focus on 'these are facts about values' and 'normative' tends to put the focus on 'these are facts about what reasons we have for action'. I will use 'evaluative' throughout this thesis for the sake of consistency.

influence of natural selection) would match the evaluative truths that realists posit. Therefore, she argues, realists are faced with the implausible sceptical conclusion that our evaluative judgments are likely mostly off track and that our system of evaluative judgments is utterly saturated and contaminated with illegitimate influence.

A common response has been to suggest that rational reflection should be able to correct for this distortionary influence, by pruning and systematising our evaluative judgments, weighing them against each other etc. we may be able to come to the evaluative truth (Street, 2006, 2008). But rational reflection is not a solution as it must always proceed from some evaluative standpoint; a starting stock of evaluative judgments that would be just as contaminated with the illegitimate influence of evolutionary forces.

> "By definition, one's starting set of views is going to be within reach of a pruned and systematized version of those very same views. So by definition, whatever method one actually used to arrive at one's starting set of views is going to have landed one within reach of a pruned and systematized version of those views. But it obviously doesn't follow from this that one's method was a good one that is likely to have landed one on the independent normative truth…" (Street, 2008, p. 216)

What this amounts to is a 'garbage in, garbage out' problem; if the starting stock of evaluative judgments is utterly contaminated by illegitimate influence, then the stock of judgments that have emerged from the process of rational reflection will also be utterly contaminated by illegitimate influence. Thus, rational reflection is not a solution to the problems associated with denying a relation between evolutionary attitudes and independent evaluative truths.

Since denying the existence of a relation between our evolutionary attitudes and the independent evaluative truths appears to be problematic, the realist might instead accept the existence of such a relation. In this case, the realist needs to give an account of the relation. Street (2006) argues that realists are forced to give the *tracking account* in response, which suggests that the presence of evaluative judgments is explained by the fact that these judgments are true, and that the capacity to discern such truths was advantageous for the purposes of survival and reproduction. However, there is an alternative account of why we make the kinds of evaluative judgments we do that Street suggests is far superior to the tracking account, the *adaptive link account*. This account suggests that certain kinds of evaluative judgments rather than others

enhanced fitness 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.

Street (2006) suggests that the *adaptive link* account holds several advantages over the *tracking account*. Firstly, it is more parsimonious in that the tracking account posits something extra that the adaptive link account does not, i.e. independent evaluative truths. Secondly, the adaptive link account is much clearer; the tracking account does not explain why it is advantageous for an organism to grasp the independent evaluative truths posited by the realist. The realist needs to give an explanation. Finally, Street argues that the adaptive link account does a much better job at illuminating the phenomenon that is to be explained. It explains why there are widespread tendencies among human beings to make some evaluative judgments rather than others. The tracking account however, can only assert that certain evaluative judgments are simply true. On the basis of these three points, Street therefore argues that it is clear that we should prefer the adaptive link account over the tracking account, and thus the realist's strategy in asserting a relation is unsuccessful.

It is clear that the EDA's of Street (2006) and Joyce (2006) share some similarities. Street's adaptive link account is essentially a version of the moral nativist hypothesis, a hypothesis that can be fully endorsed by a nihilistic moral error theorist, just applied to the evaluative domain as a whole. It provides an explanation of the formation and persistence of our evaluative beliefs, in which the truth or falsity appear to play no *explanatory role*. The tracking account is another kind of moral nativist account. However, asserting the tracking account is to assert that the truth *does,* in fact, play an explanatory role in our belief-formation processes, that the moral error theorist *cannot* endorse the genealogical hypothesis. But then the challenge is to explain how this is the case, why it is advantageous to track the truth in this circumstance, and why certain evaluative judgements are true rather than others; no easy task. If Street is correct that the adaptive link account is superior, then we have no reason to think that the truth does play an explanatory role in our belief-formation process. Street's EDA is therefore also a burden shifting argument, placing the burden on the realist to provide an account of why and how the tracking account is superior.

That being said, Berker (2014) argues that the argument for why realists must accept the tracking account if they take the second horn of the dilemma equivocates between the tracking account in a broad sense and the tracking account in the narrow sense.

> **Tracking account (in the broad sense):** Evolutionary forces have tended to make our [evaluative] judgments track the attitude-independent [evaluative] truth.
>
> **Tracking account (in the narrow sense):** Evolutionary forces have tended to make our [evaluative] judgments track the attitude-independent [evaluative] truth *because* it promoted our ancestors' reproductive success to make true [evaluative] (proto) judgments (Berker, 2014; pp: 13-14)

As can be seen, the narrow tracking account is a specific version of the broad tracking account. Berker (2014) argues that if Street (2006) is arguing for the tracking account in the broad sense, it is not clear that such an account is scientifically unacceptable, only the narrow sense would be. This is because, while Street would be correct that the realist who asserts that there is a relation between our evaluative judgments and the attitude-independent evaluative truth must accept the tracking account in the broad sense, the narrow tracking account is not the only way of satisfying the tracking account in the broad sense. There are many possible stories that are compatible with the broad tracking account but explain differently exactly how the evolutionary forces tended to make our evaluative judgments track the independent evaluative truth. So it would not matter that the narrow tracking account is not acceptable so long as a different version of the broad tracking account is.

One way of satisfying the broad tracking account without recourse to the narrow tracking account is through third-factor accounts, including moral naturalism (See section 1.1). Because the relevant relationship involved in such third-factor accounts is a grounding relation, not a causal one, they are an example of the broad tracking account without being an example of the narrow tracking account, thus potentially avoiding the problems associated with the latter. However, Street provides two major objections to third factor accounts, the 'one level up' and 'trivially question-begging' objections which we will discuss in Section 1.6. I will argue in Chapter 3 that these objections can be overcome but they do place further constraints on any successful third-factor theory.

Street (2006) finally argues that, in comparison to the realist account, the anti-realist has no problem with accepting a relationship between the content of our evaluative judgments and the content that natural selection would have tended to push us toward. Nor do they have any problem with accepting the adaptive link account or whatever explanation scientists ultimately arrive at. This

is because, for anti-realists, evaluative truth is understood as a function of the evaluative attitudes we have, however we originally came to have them. Therefore, Street suggests we should abandon the realist approach and adopt an anti-realist approach instead. In her opinion, we should adopt Humean Constructivism where evaluative truth is determined by all our evaluative attitudes in reflective equilibrium.

Street (2006) is therefore taking a revisionary approach to morality and value. She is using her EDA to undermine our justification for moral realist belief, and then attempting to provide a compelling alternative that avoids the EDA. As we can see from Joyce's (2006, 2016d) EDA, it is not only moral and evaluative realism that is targeted, but many anti-realist theories as well. For example, it has been argued previously that Street's Humean Constructivism also must face the challenge of the Darwinian Dilemma, and is targeted by her 'one level up' and 'trivially question-begging' objections (Berker, 2014; Tropman, 2014). So the anti-realist success theorist also faces a burden of proof to provide a compelling account that shows that the truth does indeed play an explanatory role in our belief-formation process. While Street does not explicitly address this point, we can see that this is what she is attempting to do by providing a positive anti-realist account that explains how our evaluative beliefs tended to track the evaluative truth: through the evaluative attitudes being the *grounds* for the evaluative truth. Whether she is successful or not is another story.

It would seem though that Joyce (2006) would argue that she is unsuccessful for the same reasons he thinks moral naturalist theories would be unsuccessful, because they fail to satisfy our most important moral desiderata, including *practical clout*, and thus fail to count as a *moral* theory at all. It is clear though, that Street (2006, 2012) would disagree that *practical clout* is a necessary desiderata for 'morality'. She even argues that the Humean Constructivist grounding account is a conceptual truth that can be reasoned to by anyone, even alien beings (Street, 2006; p. 163)[8]. Thus, the actual role of her EDA in arguing for her view is quite modest, it targets the epistemically conservative evaluative realist, pushing them to provide a compelling account in the hopes that they cannot. Meanwhile, she makes a positive case for Humean Constructivism, arguing that it avoids the EDA. The hope is that the realist, unable to make a compelling account, will be then be convinced by the argument for Humean Constructivism. It is clear then that metaethical considerations, particularly in regards to the conceptual success of revisionary approaches, play a large role in whether morality, and moral realism in particular, can be rescued from the EDA.

[8] Although Berker (2014) questions the soundness of this argument.

1.4. Moral Naturalisms: Internalism vs. Externalism

The term 'moral naturalism' does not just refer to one theory, but to a broad spectrum of theories[9]. Different moral naturalist theories may be more or less effective, and often in different ways, at resisting the challenges that Joyce (2006) and Street (2006) put to the moral naturalist. In fact, Joyce specifically divides moral naturalist theories into two camps, internalist and externalist, and deals with each separately. However, these two terms are used in a number of different ways even just within discussions of morality, so it is worth clarifying what exactly Joyce is referring to.

Moral naturalist theories can be divided up into 'internalist' theories and 'externalist' ones. Confusingly (even setting aside the fact that this terminology is used to divide and categorise theories in other domains in completely different ways), there are several different ways that one can be an 'internalist' or 'externalist' about morality. Firstly, there is the problem of the presumed *motivational* aspects of moral judgements. *Motivational internalists* insist that the motivation to act accordingly is intrinsic to a moral judgment, such that if an individual sincerely holds a moral belief or judgment they are automatically motivated to act accordingly, even if such motivation is ultimately outweighed by other concerns. *Motivational externalists*, on the other hand, deny this, arguing that sincerely holding a moral judgment and being motivated to act accordingly is merely a contingent affair; that there is no necessary connection between the two. Joyce (2006) falls into this latter camp, he argues that while moral judgments conventionally express a corresponding conative attitude which may motivate the individual, someone may sincerely hold and express a moral judgment without being motivated to act accordingly in that moment.

A different, but oftentimes related, distinction regards the presumed reason-giving nature of moral facts[10]. *Reasons internalists* argue that reasons are internal to the moral fact or judgment, such that moral facts give a reason to act accordingly regardless of anyone's desires. For example, if inflicting pain is wrong, then according to the internalist, an individual always has a reason not to inflict pain, even if that individual really enjoys inflicting pain and has no desire to act morally. This is not to say this reason cannot be outweighed by other reasons, including those derived from other moral facts, but just that there is always a *pro tanto* reason to act accordingly. *Reasons externalists* however argue that there are no-desire independent reasons for action, we only have reason to act according to the moral facts, if we desire (or perhaps desire to desire) to do so. However, most

[9] Some examples of moral naturalist theories include Cornell Realists (e.g. Boyd (1988), Sturgeon (1985), Brink (1986)) as well as others including Copp (2009), Smith, (1994) and Sterelny and Fraser (2016).
[10] This distinction was introduced by Darwall (1997).

reasons externalists would insist that we do generally desire to act morally, but this is merely a reliable contingent relationship between our desires and the prescriptions of morality. This distinction and the former are related, and sometimes confused, in that it is often assumed we will be motivated to do what we take ourselves to have reason to do, or at least we will change our motivations/desires when we realise what we actually have reason to do. However, we can conceive of a being that understands what reasons for action he has, yet has no corresponding motivation to act accordingly. Hence, if conceivability in this case is a guide to possibility, then these two dimensions are distinct.

This latter distinction, between reasons internalism and externalism, should not be confused with positions regarding Bernard Williams (1981) thesis that there are only "internal reasons". Confusingly, such a view is reasons externalist as it argues that "one has a reason to do something only if one could be motivated to do the thing by sound reasoning given one's existing motivations and given accurate non-normative beliefs" (Copp, 2012; p. 290). Such a view is reasons externalist because it insists that moral facts do not provide reasons for action on their own, such reasons are not 'internal' to the moral fact, rather an agent's reasons for action are only 'internal' to the psychology of the agent in question.

As will be seen, the distinction between reasons internalism and reasons externalism will be most relevant to this thesis. Joyce (2006) provides different arguments against moral naturalist theories on the basis of whether they are reason externalist or internalist, while Street (2008) makes clear that her EDA, the Darwinian Dilemma, targets reasons externalist and reasons internalist moral naturalism differently. As such, from now on, for the sake of expediency, I will take 'internalism' to refer to 'reasons internalist naturalism' only, and 'externalism' to refer to 'reasons externalist naturalism' only.

1.5. Joyce on Moral Naturalism

As mentioned earlier, for a moral naturalist theory to meet the challenge posed by a successful EDA it is not enough to present a merely a possible account of how we could have evolved to track the moral facts via some set of natural facts, it needs to provide a *compelling* account of how this is so. It needs to show, in detail, how the moral facts are identical to, or are grounded, by the set of natural facts picked out by the particular theory, and it needs to have some evidence that

this is the case. Furthermore, the account needs to be *believable*; we need to be able to believe that whatever properties we are talking about are *moral* properties, i.e. we need to be able to take the identity claim seriously.

> Dropping this constraint would make it the easiest thing in the world to establish that moral facts play a crucial explanatory and justifying role in our moral judgements, even assuming moral nativism. One could simply alight on any essential aspect of the proffered evolutionary genealogy—for instance, that making moral judgements improved the probability of an ancestor's genes being passed on to the next generation—and declare that that property is moral goodness (say). This reminds me of C. L. Stevenson's tongue-in-cheek example of an easy moral naturalism: 'X is morally good'='X is pink with yellow trimmings' (1937 p. 14) … By comparison, the theory that 'X is morally good'='X improves the probability of one passing on one's genes' is, while maybe not quite as silly, still well beyond the pale of being taken seriously. (Joyce, 2016a, p. 134)

Effectively, what this *Desiderata* constraint amounts to is a requirement for moral naturalist theories to satisfy some set of platitudinous desiderata for morality. Imagine we can construct a list of all the platitudes we ordinarily hold about morality. For a particular theory to be about morality, it needs to satisfy some of these platitudes. Perhaps not all of them, for we need to accommodate the fact that we can be mistaken about certain qualities of morality without it following that morality does not exist (Joyce, 2016c). If we cannot satisfy all of them, for example perhaps there are twenty platitudes of which we can satisfy 15, and no other theory can do better, then we may find satisfying fifteen is 'good enough' (Joyce, 2006; p. 191). However, conversely, there may be some platitudes of such importance that for a theory to fail to satisfy those is for that theory to fail to be talking about morality at all (Joyce, 2006; p. 191). In best case scenarios, the theory may be talking about a 'schmorality', something like morality, but different in significant ways (Joyce,2016b). In other cases, the theory may be referring to something completely different.

> Let me be clear what is meant by "schmorality" in this context. Picture a continuum comprised of what can be thought of (in a benignly vague manner) as "normative frameworks." At one end we have value systems that clearly count as moralities: Christian ethics, deontological systems, Moorean intuitionism, Platonic theories about the Form of the Good, and so on … At the other end we have things that clearly don't count as moralities: the rules of chess, etiquette, doing whatever the hell you feel like, and so on … Somewhere

> on this continuum will lie normative frameworks for which it is not immediately apparent whether they count as moralities: Some people will think they do; others will think they don't. (Joyce, 2016b; p. 55)

For Joyce (2006), what distinguishes 'moralities' from non-moral 'normative frameworks', i.e the key constraint that needs to be satisfied for such theories to count as *moral* theories, is that they must account for practical clout, or the inescapable authority of moral judgments. Joyce argues that our ordinary moral judgments are seen to hold this practical clout; moral judgements are seen to be inescapable, applying universally to everyone everywhere, and they are seen to be authoritative, providing reasons for action independent of any and all of our desires. So for any moral naturalist theory to be compelling, and thus meet the challenge of the EDA, it needs to grapple with this perceived practical clout and either find a way to fit it into the theory, or show that it is not a necessary part of the concept of morality in the first place.

1.5.1. The Reasons Internalist Approach to Practical Clout

Reasons internalist moral naturalist theories will take the former approach, attempting to locate inescapable authority in the natural world. But Joyce argues this approach is doomed to failure, as it is hard to see how any set of natural facts could provide the practical clout necessary (Joyce, 2006). Effectively, this is an instance of the 'IS-OUGHT' gap, a problem which many theorists have attempted to solve, but so far none appear to be conclusively successful. Some naturalists provide a middleman between morality and reasons, such as the utilitarian identifying moral goodness with happiness and then attempting to show that we have reason to pursue the general happiness, while others take a simpler approach where moral requirements are just whatever one has a real reason to do.

The latter approach is a version of moral naturalism known as "practical reasoning theory" which suggests that "what a person has 'sufficient reason' to do is tied to what he would want to do if he 'reasoned correctly'" (Joyce, 2006, pp. 194-195). What 'correct reasoning' is will depend on the particular account in question (examples include Korsgaard (1996) and Smith (1994)). Practical reasoning theories attempt to avoid the 'garbage-in, garbage-out' problem by extending from the domain of rationality to the domain of morality. But Joyce (2006) argues that correct reasoning is sensitive to a person's contingent desires, even if correct reasoning would suggest that an alcoholic

should refrain from drinking, it nevertheless would take his desire for drink into account. Because of this sensitivity to desires, "correct reasoning" may not always lead to the moral course of action, for example, for a man who has the desire to burn cats, "correct reasoning" may simply lead him to more effective ways of doing so, not necessarily to resist engaging in such behaviour. Thus, argues Joyce, what the moral naturalist needs is "… a substantive and naturalisable account of "correct practical reasoning" according to which any person, irrespective of her starting desires, would through such reasoning converge on certain practical conclusions that are broadly in line with what we would expect of moral requirements" (Joyce, 2006, p. 196).

To Joyce, it seems clear that such convergence is extremely unlikely, if not impossible, especially considering the amount of disagreement, both on moral grounds as well as practical (Joyce, 2016a, 2016c). On the other hand, some theories, such as Smith (1994) argue that such convergence is possible, and point to the substantial progress made in the field of ethics over the course of human history. However, Smith (1994) also notes that whether such convergence is really possible is a substantive question, and one which Smith admits we may never be able to answer. In any case, it seems to Joyce that "correct practical reasoning" is in *general* a desire sensitive affair, and thus the practical reasons it produces are desire sensitive as well. If this is the case, then "correct practical reasoning" cannot account for practical clout as the practical reasons it generates are not *inescapably authoritative* (reason giving independent of our desires) in the way that moral propositions are generally perceived.

1.5.2. The Reasons Externalist Approach to Practical Clout

The externalist moral naturalist, on the other hand, denies that practical clout is a requirement of morality at all. Instead, the moral externalist argues that whether a theory counts as a theory about morality is determined on other grounds. Each will argue that their own theory is 'good enough' or sufficient, that it satisfies enough of the other platitudinous desiderata for moral theories to count as a theory about *morality*. Whether any of these revisionary approaches to morality is 'good enough' is a controversial question, and it is possible that it may never be decided or even be decidable. Lewis (1989) contemplates similar questions in regard to his own dispositional theory of value and whether such a revision is 'good enough'.

For Lewis (1989), one of the desiderata for 'value' concerns 'non-contingency'. However, it appears that nothing that satisfies the other desiderata for 'value' also satisfies non-contingency. So genuine 'values', strictly speaking, do not exist, as they would have to satisfy an impossible set of conditions. Therefore, strictly speaking, it appears we ought to be error theorists about value. However, there are imperfect claimants that satisfy nearly everything that we want – for example, Lewis' own dispositional theory of value, where a value is just what we would desire to desire under ideal conditions. If one of these imperfect claimants is indeed sufficient, then, loosely speaking, the name 'value' may go to that claimant, and there would be values - lots of them (Joyce, 2016c; Lewis, 1989). Ultimately, Lewis argues that whether one will agree that any given imperfect claimant is indeed sufficient is a matter of temperament. Someone inclined to revolutionary outlooks or error theory will likely argue that there are no values. Others may argue that we should say what we mean, loosely or strictly. Those more conservative (like Lewis himself) may argue that there are imperfect claimants that capture almost everything that we could want out of the concept, so to continue using the term is 'good enough'.

The same could be said of morality. To avoid the EDA, the moral naturalist must come up with an account of how moral properties can be identified with, or grounded by, the natural properties. However, in order to do so, they need to show that such natural properties can satisfy our desiderata for a property to be considered 'moral'. If Joyce (2006) is right that the folk ordinarily perceive moral judgments as holding practical clout, then that is likely part of the desiderata for morality. But no moral naturalist theory appears to be able to satisfy the need for moral judgments to hold practical clout. So it may be the case that, strictly speaking, not all moral beliefs are unjustified. However, a moral naturalist theory may be able to satisfy nearly everything else we would want out of a theory about morality (including the ability to avoid the debunking force of the EDA). So the name 'moral' may go to one of those imperfect claimants, and, loosely speaking, moral beliefs would in fact be justified. But the question is then, is any theory about morality 'good enough' without practical clout? Should we become moral error theorists about morality, accepting that they are unjustified, or should we take a revisionary approach, on the proviso that there is a theory that satisfies nearly all our other desiderata?

There are some clear cases where revisionary approaches have been rejected or simply not taken. Consider the term 'witch' (as an example taken from Joyce (2006, 2016b)), which refers to a 'woman with supernatural powers'. When it became clear that belief in supernatural powers was

unjustified[11], we likewise adopted an error theory about witches, plainly that they do not actually exist. It is possible that we could have revised our witch discourse, identifying the term with women who play a certain role in the community, or with whoever is designated a witch by their community, but such approaches were either never taken up or rejected. However, this is a clear case where revision was not 'good enough'. Both 'value' and 'morality', on the other hand, fall into a grey area where there is much disagreement, and it is hard to see how we could resolve such disagreement.

While Joyce (2006, 2016b, 2016c) argues that it is hard to see how the debate may ever be settled, he suggests that we might have a decent chance of finding out if we give consideration to what the concept is used for, whether the revised discourse can continue to carry out its role. Joyce (2016b) brings up the concept of 'simultaneity', which, after Einsteinian theory showed that nothing was strictly speaking simultaneous with anything else, could safely be revised to mean 'relative simultaneity' with little impact on everyday use. In regards to witch discourse, Joyce (2016b) argues that the concept 'witch' was not adjusted to any of the possible revisionary approaches (including those mentioned above) because none of them would allow people to put the term to the same use as before: "…to condemn these women for their evil magical influence and justify their being killed" (Joyce, 2016b; p. 56). However, it should be noted that the term actually *is* still in use as a way to denigrate women, just in an ironic sense. People who use the term 'witch' in this way to denigrate women do not actually believe that their targets have magic powers, but the point is still to condemn women for 'evil influence' and justify attacks towards them. It would appear that the term has been revised and put to the same use. In these ways discussed, function of a term could therefore partially explain whether the concept can be revised.

In trying to determine whether an revision of the concept of morality is 'good enough', Joyce (2006) considers the question of whether moral discourse can continue to fulfil its function on such an externalist precisification, where there is only, at most, a reliable contingent relationship between its prescriptions and people's reasons for action. He contends that it cannot. Joyce argues that the reason we think and talk in moral terms is that doing so acts as a bulwark against weakness of will, doing a better job than just deliberating about what is desired and how it might be achieved (Joyce, 2006; p. 109). But if moral prescriptions are only contingently connected with reasons for action, then there may be cases where someone ought to do something even though it is wrong to do so, simply because they have desires and/or reasons to do so.

[11] This could be either in terms of a loss of justification for witch beliefs, or in terms of a realisation that such beliefs were unjustified all along. Neither approach affects the point under discussion.

For example, a serial murderer who loves murdering may, on this account, admit to murder being wrong but find they ought to keep murdering people because they have reasons to do so (they have the desire) and the fact that murder is wrong does not give them any inherent reason to not do it. In such a case, thinking in moral terms on an externalist precisification fails to act as a bulwark against weakness of the will for the serial murderer (Joyce, 2006; p. 203). Furthermore, having an appreciation of the desire-contingent nature of moral reasons, may lead an individual to be tempted to change their desires away from acting morally if the desire is weak enough (Joyce, 2006; p. 206). Hence, Joyce concludes that in order to fulfil their function, moral propositions must be seen to have *inescapable authority*. Moral naturalism without clout therefore has no access to this reason for morality; the answer to the question of "why do we need a distinct moral discourse?" becomes "we do not". Morality becomes, in effect, much like etiquette; it holds no sway on the behaviour of those who do not 'buy into' the institution.

In summary, according to Joyce (2006):

> Moral naturalism without clout, first of all, seems to enfeeble our capacity to morally criticize wrongdoers; second, it might actually encourage wrongdoing for certain persons; and third, it renders moral language and moral thinking entirely redundant. (Joyce, 2006, p. 208).

Therefore, according to Joyce (2006), such a value system lacking in practical clout could not effectively play the social roles to which we put morality. Practical clout is thereby considered a necessary desideratum for the concept of morality.

In Chapter 2, I will be making the argument that, in actual fact, so long as Joyce's revisionary version of moral fictionalism is possible, then morality on an externalist precisfication should be able to fulfil its functional role. If my argument is successful, then it should show that this 'argument from function' is not enough to show that externalist theories cannot be 'good enough', and therefore it should show that this argument against the likelihood of a compelling moral naturalist account that can meet the challenge of the EDA is unsuccessful.

1.6. Street on Moral Naturalism

Street (2006; 2008) puts forward objections to moral naturalism by way of providing objections to value naturalism. Berker (2014) argues that there are two main objections to third factor approaches, including value naturalism, in Street's work. The first is that these approaches only put off the Darwinian Dilemma to a higher level, such that the Darwinian Dilemma can be run 'one level up'. The second is the claim that third-factor accounts are 'trivially question-begging', that they must appeal to substantive moral/evaluative truths in order to explain how we were selected to track those truths, begging the question at hand. There is a third argument that Street (2008) can be read as bringing to bear against externalist moral naturalist theories (or at least against that of David Copp (2008)). This latter argument is also concerned with the desiderata for the concept of morality, essentially arguing that externalist approaches fail to count as realist theories of *morality*.

1.6.1. The 'One Level Up' Objection

Berker (2014; p. 18) gives a good single-line summary of the third-factor theorist's central claim:

(G) Non-normative fact F (at least partially) grounds normative fact N

In order for the third-factor theorist to count as genuinely realist on Street's (2006) taxonomy they must believe that (G)'s truth does not depend on any facts about our evaluative attitudes. But then the question arises as to how does one know that (G) is true? For example, if the evaluative facts are identical to some set of natural facts, how do we know this is the case, and furthermore, how do we know *which* natural facts the evaluative facts are identical to? That is to say, how do we know what are the correct natural-evaluative identities? Street argues that the most common answer by value naturalists is to proceed roughly as we currently do proceed. She quotes Sturgeon (1985) saying "if a full account of which natural facts evaluative facts are identical with is to be had, then this account "will have to be derived from our best moral theory together with our best theory of the rest of the world"" (Street, 2006; p. 139). Essentially, we have to rely on substantive moral theory, theory which is thoroughly saturated with evolutionarily influenced evaluative attitudes.

> [The naturalist approach is to start] with our existing fund of evaluative judgments, giving more weight to those evaluative judgments which strike us as correct if anything is (for instance, the judgment that Hitler was morally depraved), and then working to bring our evaluative judgments into the greatest possible coherence with each other and with our best scientific picture of the rest of the world. (Street, 2006; pp. 139-140)

In effect, in order to determine the correct natural-evaluative identities, or to determine (G), we need to make use of our evaluative attitudes and intuitions. And we have no reason to think that these evaluative attitudes are any less influenced by evolutionary forces than our ordinary evaluative judgments. Thus, we have no reason to think that our judgements that (G) is true are no less influenced by evolutionary forces, no less targeted by the Darwinian Dilemma, than our first-level evaluative judgments.

Since the evaluative attitudes used in determining (G) (or the correct natural-evaluative identities) appear to be largely influenced by evolutionary forces, Street (2006) raises the following question: what, according to the realist who adopts a third-factor approach, is the relation between the evolutionary forces that have influenced our judgment that (G) is true (or our judgments about the correct natural-evaluative identities) and the attitude independent fact that (G) is true (or the facts about the correct natural-evaluative identities)? The realist can once again either assert or deny that there is a relation between the two. If they deny a relation, then Street argues that it would be an implausibly large coincidence that (G) would be true, and thus we are likely wrong to think so. If the realist asserts a relation, arguing that evolutionary forces pushed us towards the truth of (G), then Street would argue the realist once again must accept a tracking account, which once again loses out to the more scientifically acceptable adaptive link account. In fact, Street argues that it is hard to see how tracking something as esoteric as independent facts about what grounds evaluative facts could have ever promoted reproductive success in our ancestor's environment. In effect, the third-factor theorist is once again caught in the Darwinian Dilemma, just 'one level up' this time. The burden of proof is shifted onto the third-factor theorist to show that the judgements behind (G) are reliable.

Many theorists attempt to defend moral naturalism from Street's (2006) Darwinian Dilemma while ignoring that it does not just target moral realism, but evaluative realism as a whole. They will often attempt to derive what is morally right or wrong from what humans value or need. For example, Copp (2008) grounds moral facts in facts about what sorts of rules would best allow a

society to meet its needs, and Fitzpatrick (2014) argues that we can ignore the EDA because our moral beliefs are not wholly the result of evolutionary influence but many come from having grasped the moral facts through "our ongoing experience of various forms of value and informed reflection on it whereby we come to understand, for example, that certain features of actions are wrong-making in light of those values" (p. 246). By ignoring the original target of the Darwinian Dilemma, these theories are therefore subject to a kind of 'one level up' dilemma where it is not the moral beliefs themselves that are subject to the EDA, but our evaluative attitudes that we rely on to track the evaluative facts that serve as the grounds for the moral facts. The moral beliefs only become a reliable guide to the moral facts so long as our evaluative attitudes are a reliable guide to the evaluative facts, and it is exactly the reliability of our evaluative attitudes that Street's EDA seeks to question. Moral naturalist theories therefore may need to grapple with this EDA 'one level up' from what they first had assumed. In Chapter 3, this is one of many questions I will discuss and attempt to answer.

1.6.2. The 'Trivially Question-Begging' Objection

Street's second objection to third factor accounts is most clearly expressed in her reply to Copp (2008), and is essentially that such replies are 'trivially question-begging' (Street, 2008), that is they must assume the truth of substantive moral and evaluative theory in order to prove that such theories are reliable.

> "It is no answer to this challenge simply to assume a large swath of substantive views on how we have reason to live … and then note that these are the very views evolutionary forces pushed us toward. Such an account merely trivially reasserts the coincidence between the independent normative truth and what the evolutionary causes pushed us to think; it does nothing to explain that coincidence." (Street, 2008; p. 214)

A common objection to the 'trivially question-begging' argument is the "companions in guilt" argument (Street, 2008, p. 216). The "companions in guilt" argument is that if we are unable to appeal to substantive theory in order to establish the reliability of our evaluative/moral knowledge, then we likewise would be unable to appeal to substantive theory in order to establish the reliability of knowledge in other domains, such as science or perception. Such a ban, rather than simply putting our knowledge of our reasons for action at risk, would lead to universal scepticism on the

whole. Since we do take it that we have knowledge in these other areas, that is we can resist the 'trivially question-begging' argument in these other areas, we can likewise resist this argument in the evaluative domain.

However, the 'companions in guilt' argument can be resisted by recognising that the 'trivially question-begging' objection is only levelled against attempts to rescue justification in response to the EDA; the argument is not questioning knowledge in domains in which we are entitled to be epistemic conservatives. As Joyce (2016a, 2016d) argues, the EDA targets the epistemic conservative who argues that firmly held beliefs are 'innocent until proven guilty ', Street's (2006) EDA is no different in this regard. The EDA uses an empirical, evolutionary argument to target the epistemic conservative about moral/evaluative beliefs. That same argument however does not target the epistemic conservative about other domains of knowledge, perceptual beliefs for example, as the tracking account is simply not a controversial option in those cases. The moral/evaluative success theorist however, needs to provide some account that satisfies the broad tracking account, and this is when the trivially question-begging argument applies. The account provided by the success theorist cannot simply assume the truth of substantive moral and evaluative theory in order to prove such theories are reliable, to do so is to be an epistemic conservative, a position targeted by the original EDA. What this amounts to is the need for a compelling account that meets the *Good-Reason* constraint; we need actual evidence (or reasoning out from non-evaluative facts) that the broad tracking account is satisfied, not the mere possibility that it is. Justification for our beliefs about supernatural facts such as ghosts is not rescued simply because '*if* ghosts existed, then it would explain our beliefs about them' is true; actual evidence that ghosts exist is needed (Joyce, 2006; p. 189). Other domains of knowledge, by contrast, do not require a compelling account for them to be made, at least not in response to an EDA, because the EDA never targeted them in the first place. Essentially, in such domains we are entitled to be epistemic conservatives; we never had reason to think our justification for such beliefs was undermined, so we never had reason to establish a compelling account for such beliefs.

Another way of looking at this is by reference to Vavova's (2014, 2015) discussion of the two versions of how the EDA is commonly put. The more successful version proceeds from the empirical evolutionary premise to show you that you have good reason to think your moral/evaluative beliefs are unjustified. The less successful one, similar to other general sceptical arguments, tries to show you that you have no good reason to think your beliefs are justified. The trivially-question begging objection is a version of the latter argument (Vavova, 2014); it only seeks to show that since you

have no good reason to think your beliefs are justified, you cannot use them to show that your beliefs are justified. This latter argument does not have much effect on the epistemic conservative; they can deal with it in ways similar to how they deal with other more general sceptical arguments. The epistemic conservative however cannot so easily dismiss the former argument; they are faced with actual evidence that their beliefs *are* 'guilty', and not so 'innocent'. It is only then that the trivially question-begging objection applies. If the success theorist brings up the possibility that the moral/evaluative beliefs are identical to the natural facts as a way of showing that we could have evolved to track them, the objection argues that the success theorist is 'begging the question', assuming exactly the reliability that is put under question by the EDA. The role of the objection is to reinforce the need provide evidence to overcome the evidence presented by the EDA.[12]

1.6.3. Street and Moral Desiderata

In certain sections of her reply to Copp (2008), Street (2008) could be read as implicitly arguing that externalist moral naturalist theories fail to fully satisfy the desiderata for being realist theories of *morality*. On this reading, similar to Joyce (2006), her argument is that externalist theories fail to appropriately vindicate the objective bindingness of morality (i.e. that they fail to vindicate practical clout). Street's (2008) paper is not entirely clear though, and in later works (for example, Street 2012), Street seems to take the view that the analytic definition of morality need

[12] The upshot of this is that Street's (2006) Humean Constructivism is not immune to the challenge posed by the trivially question-begging argument. Her Darwinian Dilemma does not only target evaluatively realist beliefs as she claims, but all evaluative beliefs. It is just that Street (2006) argues that Humean Constructivism (and other anti-realist views) can accept a version of the broad tracking account that can endorse the adaptive link account. Berker (2104) sums up Street's (2008a) view of Humean Constructivism as follows:

> (G'): The non-normative fact [A judges <I have conclusive reason to *φ*>, and her *φ*-ing does not conflict with anything else she more deeply judges that she has reason to do] grounds the normative fact [A has conclusive reason to *φ*] (Berker, 2014; p. 20)

Berker (2014) argues that this anti-realist grounding account bears some remarkable similarities with the third-factor account, their core strategies for resisting the Darwinian Dilemma are the same. They both attempt to bridge the gap between the normative and non-normative realms by appealing to some sort of grounding relation. For the third-factor account this is done by grounding moral facts in some set of natural facts, while for the Humean Constructivist, this is done by grounding the evaluative facts in our evaluative attitudes, such that evaluative attitudes tracked the truth because they are the grounds for their own truth. But Street cannot simply raise the possibility of this anti-realist grounding account, for then her own view would be subject to the trivially question-begging argument. Instead, she makes the claim that the grounding claim involved in Humean Constructivism is a conceptual truth, not an evaluative one. The aim here is to show that Humean Constructivism is not merely possible, but is actually the case, and that anyone can reason to it using only conceptual truths. Of course it is controversial whether she is successful, but the point is that Street also must provide a compelling case for justification of our evaluative beliefs to be re-established.

not include morality being objectively binding independent of any of our evaluative attitudes. For the sake of argument, I will assume that this reading of Street (2008) is correct; considering its similarity to Joyce's argument, it should be taken seriously on its own merits.

Copp (2008) attempts to show that his version of moral naturalist realism meets the challenge posed by a version of Street's (2006) Darwinian Dilemma that only targets morality[13]. Perhaps because the original Darwinian Dilemma targets the evaluative domain, in her reply, Street (2008) seems to take Copp as attempting to vindicate evaluative realism as well, and so argues that either his strategy does not succeed, or his theory is not a version of realism about *normativity*[14].

Street's (2008) primary issue with Copp's (2008) account is that it appears to provide no reason *simpliciter* for action independent of any of our evaluative attitudes. She argue that, on Copp's view, there are moral reasons and non-moral reasons (for example self-grounded reasons), but no reason *simpliciter* for action, even though it purports to be a version of realism about *normativity*[15]. Furthermore, if the moral reasons and non-moral reasons conflict, there is no answer

[13] He claims that changing the target from the evaluative domain to the moral domain does not have an effect on the force of her argument. I argue in Chapter 3 that ignoring the original target of the evaluative domain does in fact change the force of her argument.

[14] This may also be an example of Copp (2008) and Street (2008) talking past each other, perhaps due to differing usage of the word 'normativity'. While Copp's (2009) pluralist-teleological theory may purport to be a realist theory of normativity, it is not evaluatively realist in the sense meant by Street (2006, 2008). In fact, at the end of Copp (2009), he attempts to avoid contention about the use of the word 'normativity' by admitting:

> Unfortunately, however, there is little agreement among philosophers about how to use the term "normativity," and it can often seem that philosophers who discuss normativity are talking past one another. Let me therefore set aside the word. I hope to have at least shown that pluralist-teleology provides a unified account of the truth conditions of a class of judgments that bear on solutions to practical problems that are endemic to the human condition. (Copp, 2009; p. 36)

Street (2008), on the other hand, seems to gather from Copp's other works that he intends his view to be normatively realist in the sense targeted by the Darwinian Dilemma. This is certainly understandable, given Copp's (2008) talk of reasons and normativity, but as seen from the quote above, he is really discussing normativity in a very different sense. I do not want to go into this discussion in much more detail here, but I thought it important to raise the difference in terminology. The more relevant discussion is how Street (2008) generalises from her discussion of Copp (2008) on an externalist reading.

[15] Copp (2009) attempts to deal with the reasons *simpliciter* objection by claiming that self-grounded reasons have 'default priority' in evaluating deliberation because such reasons are always relevant to evaluating deliberation, given what it is to deliberate. He therefore argues that "the default is to interpret the 'ought simpliciter' as the ought of practical rationality" (Copp, 2009; p. 36). Therefore, what reason we have to be moral, to endorse the moral system, will be derived from our self-grounded reasons. However, this still does not avoid with Street's (2008) objection that such a theory is not *realist*, since one's reasons for acting morally will still be contingent on one's evaluative attitudes. Street (2008) argues that Copp's (2008) view "doesn't construe morality as objectively binding in the way one might have thought a realist theory aspires to, or indeed in any way that wouldn't be perfectly acceptable to an antirealist about normativity, who holds that things are required ultimately because we take them to be" (p. 211).

as to which you should act according to; there is no answer to the question 'what should I do period'. In response to this, Street (2008) says "… this sounds like an error theory rather than rather a version of moral realism: are you really telling your child that whenever he (or anyone else) asks what he should do *period*, the question is confused?" (p. 221).

Against the objection that Copp's (2008) theory still provides guidance because it tells us what we have *moral* reason to do, how *morally* we should live, she argues:

> Assuming it takes no stand on how to live period, the theory has no more normative implications than does an analysis of the function of Jim Crow laws or the rules of tiddlywinks. One could equally well say of these analyses that they tell us what we have *Jim Crow* or *tiddlywinks* reason to do—that they tell us how in a *Jim Crow* or *tiddlywinks* way to live. Such theories provide "guidance" only in a trivial sense that is analogous to the sense in which a descriptive statement of means to an end provides guidance. (Street, 2008; pp. 221-222)

Joyce (2006) makes a similar argument regarding externalist theories advocating a 'moral' system that appears more like etiquette than a true morality. Street (2008) can therefore be read as also arguing that an externalist theory ends up being a realist theory about a 'schmorality' (Joyce, 2016b) rather than a true 'morality'. Street (2008) argues that a version of naturalist realism that fails to meet the criterion of being reasons internalist "is perhaps realist, but not normative realist: it vindicates the objective bindingness of morality in just the same way that an analysis of the function of Jim Crow laws vindicates the objective bindingness of segregation, which is to say not at all" (Street, 2008; p. 224). So Street appears to be arguing that the ordinary conception of morality is an internalist one, where moral facts are objectively binding independent of anyone's evaluative attitudes.

The externalist moral naturalist attempts to revise the concept of morality to abandon practical clout, abandoning evaluative realism. However, Street (2008) argues that to abandon reasons internalism, to abandon practical clout, is:

Copp (2009) does not seem unamenable to his theory being antirealist about normativity in this sense, at one point calling his pluralist-teleological view a "'constructivist' picture" (Copp, 2009; p. 23). However, he still would insist his theory is *morally* realist, as well as *normatively* realist in his own sense of the term. It therefore may very well be that Copp's externalist moral naturalist theory and Street's (2012) Humean Constructivism are in fact compatible.

> …to abandon [evaluative] realism in any sense that *vindicates* morality. It is to compromise, and to acknowledge (either unintentionally or in a less than fully upfront way) the exact same conclusion that antirealists argue for, namely that there are no genuinely normative facts or truths that hold independently of all our evaluative attitudes. (Street, 2008; p. 224).

According to Street's view then, externalist moral naturalism is really an anti-realist perspective. This is because what is really at stake in the realist/antirealist debate in metaethics, according to Street, is whether there are facts about what reasons (*simplicter)* we have that are independent of any of our evaluative attitudes. However, this is not a description of the realist/antirealist divide that is universally held. Joyce (2016a) gives a definition of moral realism (following Handfield (2016)) which consists of the following:

> … two semantic claims—(i) that moral discourse should be interpreted literally and (ii) that it is truth-apt—plus two substantive claims—(iii) that at least some of the discourse is true and (iv) that this truth is mind-independent. (Joyce 2016a; p. 127)

While it may be the case that much of the ordinary moral discourse assumes that uncompromising normative realism holds, moral realism on this definition only holds that some of the discourse is true and mind-independent, which is perfectly coherent with normative realism, as Street (2008) understands it, being false. For example, it may be that there are indeed moral facts that are mind-independent but the reasons to act accordingly are dependent on ones' desires to act accordingly. Berker (2014) argues that according to Street's (2006, 2008) definition of evaluative realism, many ethical theories we ordinarily take to be compatible with realism, such as preference utilitarianism may turn out to be antirealist theories. This conflict in definition is part of the well-known difficulty regarding how to situate constructivist theories such as Street's in relation to realism and anti-realism (Bagnoli, 2017). And in some ways, Street appears to recognise this difficulty; as mentioned, she acknowledges that naturalist theories may be realist about *something*, even if they are not realist about normativity in her sense. The question is then, are these sorts of realist externalist naturalist accounts, *moral* realist accounts? That is to say, are they realist theories about *morality*, rather than realist theories about some *schmorality*? Both Joyce (2006) and Street (2008) seem liable to answer in the negative, arguing that moral realist theories need to account for practical clout in order to be realist theories about morality at all.

The question of whether externalist moral naturalism counts as a realist theory of morality therefore appears to boil down to whether the theory in question is 'good enough' to count as a theory of morality at all. If we can show that *something*, that schmorality, can claim the title 'morality', then externalist naturalist theories will be realist theories about morality, even if they do not vindicate the objective bindingness of morality.

Conclusion

In this chapter, I have reviewed Evolutionary Debunking Arguments (EDAs) for moral realism in general and outlined two major formulations of the EDA, those of Joyce (2006) and Street (2006). In particular, I showed how the conclusion of the EDA is epistemological in nature, undermining our justification in our moral beliefs and placing the burden on the moral success theorist to provide a compelling account of how we evolved to track the moral facts. In particular the moral success theorist must provide an account that meets two constraints to be compelling: the *Good-Reason* constraint and the *Desiderata* constraint. I also outlined how Joyce and Street argue against the possibility of such a compelling account being existent or forthcoming. In the next two chapters I will discuss how at least some forms of moral naturalism may be able overcome at least some of these challenges. In Chapter 2, I will discuss how externalist moral naturalism might overcome Joyce's objections, introducing a new variety of externalist moral naturalism to do so, and in Chapter 3 I will discuss whether moral naturalism, and that new variety of externalism specifically, can overcome Street's objections to the value naturalist. If I am successful, then I may be able to show that we can avoid the debunking force of the EDA against morality.

Chapter 2:

Moral Naturalism and Moral Fictionalism as responses to Joyce's EDA

In this chapter, I will focus on the adequacy of moral naturalism, specifically externalist moral naturalism, as a response to the Evolutionary Debunking Argument (EDA) outlined by Joyce (2006). As discussed in Chapter 1, the EDA presents an epistemological challenge that shows our moral beliefs are unjustified. To meet this challenge and rescue justification, a *compelling* third-factor account (for example a moral naturalist theory) that shows how the moral truth plays an explanatory role in our belief formation processes must be provided. However, Joyce (2006) argues that no such compelling moral naturalist account is extant or forthcoming, because no moral naturalist account can appropriately satisfy our desiderata for a theory to be about *morality*. In particular, naturalist theories appear unable to account for the perceived practical clout of morality. Joyce argues that people perceive moral judgments as holding practical clout, as being *inescapably authoritative*. That is to say people see moral propositions and judgments as being *inescapable*, applicable universally to everyone without exception, and as being *authoritative* or *reasons internalist*, providing reasons on their own to comply with the moral proposition in question, independent of any of our desires. It would appear, however, that naturalistic facts are unable to hold these properties. Joyce argues that revisionary naturalist approaches that abandon practical clout are unsatisfactory as theories about *morality* as the 'moral' discourse implied by such approaches would be unable to continue to play the same role as the old discourse, that of regulating our behaviour and promoting cooperation. Joyce therefore argues that it is likely impossible that moral naturalist approaches could ever meet the challenge posed by the EDA and rescue our justification in our moral beliefs.[16]

In this chapter, I will use an argument Joyce (2005, 2006) makes for moral fictionalism to show that, if moral fictionalism can be successful, then a revisionary externalist moral naturalist discourse could, in fact, be successful as a *moral* theory as well. Joyce (2006) suggests that in response to the EDA undermining our justification in morality we should not become moral abolitionists, but rather, that we should treat morality as a "useful fiction". To make use of morality, Joyce (2005, 2006) argues, we need not believe in it or the rules it supplies, rather we need merely to accept and follow its rules. Thus, Joyce denies the idea that getting regulative benefits from

[16] For a recap of Joyce's (2006) argument against the possibility of a compelling moral naturalist theory in standard form, see Appendix B.

morality actually requires belief in its propositions as literally true. But, as mentioned, he also says that moral judgments must be taken to be authoritative or internalist in order to provide their regulative benefits. I will argue that if obtaining the regulative benefits of moral discourse is possible when there are no moral truths (through make-believing that there are), then obtaining the regulative benefits must be possible when there are moral truths, be they internalist or externalist. However, if the moral truths are externalist, then it may be the case that to obtain the full benefits of morality we may need to make-believe that such truths are in fact internalist in nature.

In the first part of this chapter, I will outline Joyce's (2005, 2006) argument for moral fictionalism in more detail, starting by outlining the value of morality, that is to say the three major benefits that morality provides; the benefits relating to personal commitments, dyadic commitments and social coordination. In other words, how morality can act as a 'bulwark against weakness of the will', how it can promote cooperation in small groups by acting as a 'commitment device', and how morality can promote 'correlated interaction' by linking self-directed and other-directed moral judgments, respectively. I will then show how moral fictionalism is able to achieve these three types of benefit by utilising a 'precommitment' to morality and moral discourse.

However, there are a number of concerns regarding fictionalism that give us some reason to limit what we make-believe. If we can achieve the benefits of moral discourse by believing only what is true and without having to maintain an extensive fiction, then that seems, on the face of it, the preferable option. I will argue then that if it is possible that the moral fictionalist can achieve the benefits of morality, then the externalist moral naturalist, in taking a revisionist approach, can also acquire at least the first two benefits by relying on similar psychological mechanisms. I further argue that a revisionist and a highly effective moral fictionalist will likely end up with very similar set of rules, but the revisionist does not need to rely on a fictionalist attitude. It would therefore appear that if a moral fiction can fulfil the role of the original moral discourse, then a revised externalist moral discourse should be able to play at least some of the same role as well. By Joyce's (2006) lights, this would make it less clear that externalist naturalist revision of morality would not be 'good enough'.

In the second part of the chapter, I will show how a variation of externalist moral naturalism can achieve the third major benefit of morality, the social benefit, that of linking self-directed and other-directed moral judgments. I will begin by showing what it is about ordinary moral discourse that can provide us this benefit, namely the perceived objective bindingness of morality. Thus, I will

reveal why the standard externalist moral naturalist, who explicitly rejects the objective bindingness of morality, is unable to achieve this third benefit of morality, while the fictionalist, who makes believe in such objective bindingness, actually can. I will argue though, that the externalist moral naturalist can adopt part of the fictionalist strategy while not accepting all of it. They can insist, while in their most critical contexts, on an externalist naturalist approach to moral discourse, but in their ordinary contexts, they can make-believe that moral facts are objectively binding independent of their desires. I call this the fictional-internalist externalist, or FI-externalist for short. I argue that the FI-externalist would have access to all the benefits the ordinary fictionalist can acquire whilst limiting the barely-stable fiction as much as possible. If the determination of what revisions of morality are 'good enough' is based on whether the revised discourse can play the same role as the original discourse, then it would seem that a FI-externalist theory should acceptable, provided the theory is otherwise compelling (satisfying the Good-Reason constraint for example).

2.1. Moral Fictionalism[17]

Joyce (2006) lays out his argument for why moral discourse is assertoric, i.e. that moral judgments express belief states, and that moral assertions are typically untrue or at least unjustified. In addition, on the basis of these two claims, Joyce (2006) presents the outline for an argument for moral fictionalism. However, in an earlier paper (Joyce, 2005), he presents a more detailed version of this argument which I will draw on here. Joyce (2005) draws from an analogy to fictionalism about colour, wherein the colour fictionalist (named 'David') adopts an error theory about colour but continues to use colour discourse in 99% of his life. It is in that last 1% though, which Joyce suggests are those contexts where David is at his most "undistracted, reflective and critical", for example the philosophy classroom, where David will espouse his error theory about colour. It is his pronouncements in those most critical contexts that reveal David's true beliefs. The rest of the time it is not that he does not hold these beliefs, it is just that he is not attending to them. It is important to note that Joyce does not believe that it is enough that *if* David were in his most critical context *then* he would espouse an error theory of colour, rather David must have actually inhabited this most critical context at some point in his past and be disposed to do so if placed in his most critical context in the future. As long as this is so, then David, according to Joyce, actually believes that the world is not coloured at all times. Joyce argues that we can do the same thing with morality; as long as we have denied the existence of morality when in our most critical contexts and continue to be

[17] For a summary in standard form of Joyce's (2005) argument for moral fictionalism, see Appendix B.

disposed to do so, we can adopt an error theory about morality but still utilise moral discourse in 99% of our lives without contradiction.

On the question of whether to adopt moral abolitionism or moral fictionalism after accepting a moral error theory, Joyce (2005) argues that we need to consider what the value of morality is and whether moral fictionalism can capture this value, at least to some degree. We need to understand *why* morality might have evolved. Joyce (2005; 2006) puts forward that the main benefit to the moral sense is its ability to support inter- and intrapersonal commitments and link them together, that the benefit of morality lies in its ability to promote cooperation through the regulation of behaviour and its ability to change the incentive structure behind cooperative enterprises.

2.1.1. The Value of Morality

Joyce (2005, 2006) lays out several likely benefits that morality evolved to bestow; the benefits relating to personal commitments, dyadic commitments and social coordination. While Joyce (2006) discusses all three benefits, Joyce (2005) focuses on the role morality plays in supporting personal commitments, providing prudential benefits from cooperation. He argues that on average it is in one's best interest to act cooperatively with others, but people often face competing desires, e.g. the pursuit of short-term profit. While it is possible that in some situations, the pursuit of a short-term profit will benefit the agent in the long-term, due to the fact that humans are imperfect reasoners, they will often mistake situations where acting uncooperatively will *disadvantage* them long-term for situations that will *benefit* them long-term. So it is rational to adopt a rule to always act cooperatively when others do as well. However, as mentioned, people are imperfect reasoners and will not always act rationally, especially when they are tempted by short-term profits. Morality, argues Joyce, acts as a bulwark against this weakness of will.

> The hypothesis, then, at its first approximation, is that a judgment like "that wouldn't be right; it would be reprehensible for me to do that" can play a dynamic role in deliberation, emotion and desire-formation, prompting and strengthening certain desires and blocking certain considerations from even arising in practical deliberation, thus increasing the likelihood that certain adaptive social behaviours will be performed (Joyce, 2006; pp. 113-114).

This view of morality can be likened to Dennet's (1995) idea that moral judgments act as "conversation-stoppers" that prevent further deliberation on a course of action, thereby preventing temptation towards uncooperative acts that are profitable in the short term. However, morality need not ensure cooperative behaviour in order to be adaptive, it need only increase the likelihood that one will cooperate (given likely reciprocation) on average (Joyce, 2006).

Joyce (2006) also highlights the key role of the specifically moral emotion of guilt in supporting personal commitments as evidence that mere non-moral emotion could not play the same role in regulating behaviour. This is because the feeling of guilt requires "the thought that one has transgressed against a norm" (Joyce, 2006; p. 112) and involves a belief that one deserves punishment or must make amends. So while one may be motivated through sympathy to avoid harming others, or to make amends after causing harm, this is not as robust as guilt and fades over time. In such a case there is no resolution that something *must* be done, and an individual motivated merely by sympathy may, upon committing harm, actually be motivated to distance themselves and not think about what they have done instead of seeking to address it (Joyce, 2006). If this is the case, the moralised thinker has an edge over the non-moralised thinker; they can have all the same sympathies and inclinations as the non-moralised thinker but have access to this more robust form of self-recrimination, one that ought to make their motivation to avoid harm more resolute than otherwise. This support for personal commitments is the benefit of morality that Joyce (2005) seeks to capture through establishment of a fictionalist alternative to abolitionism. However, in his book (Joyce, 2006) he also discusses the benefits accrued to one's interpersonal commitments.

While Joyce (2006) argues that morality might be advantageous through its ability to support prudent actions, he draws on Frank (1988) to argue that in many cases, strengthening the likelihood of even imprudent actions may be beneficial, especially in regards to cooperative interactions. The idea is that one can derive benefits from holding and signalling an emotional commitment towards some imprudent action in certain circumstances. For example, if one gets indignantly angry about being sold a faulty item, even if it is only ten dollars, and one has a reputation for such, then such an individual is less likely to get cheated by a shopkeeper than an individual who has a reputation for prudent action. Even though the effort spent in seeking a return may not be worth the ten dollar gain, the act of doing so gives two other benefits. First, it strengthens the commitment, making it more likely the individual will seek redress in other situations. Second, the costly act builds a reputation and signals to others that the individual will go to great lengths to seek redress in such

situations, making it less likely the individual will be cheated (it is more costly to cheat them) and thus making it less likely the individual will actually have to engage in the imprudent behaviour. The latter benefit is the 'primary value' of the commitment, that is the benefit derived from communicating a willingness to do something, whilst the former benefit is the 'secondary value', that is the benefit derived from actually doing it (Joyce, 2006). It can therefore be seen that so-called imprudent actions undertaken as a result of such a commitment device are not truly imprudent; the primary and secondary value associated with the commitment changes the incentive structure of the interaction, making seeking the redress of the ten dollars the most prudent, and rational, thing to do.

Joyce (2006) and Frank (1987, 1988) suggest that moral conscience is another such commitment device. However, while Frank argues for the central role of emotions as the commitment device and difficult-to-copy emotional displays as the signalling device, Joyce argues that language and belief also play an important role in moral conscience and its display. For instance, Frank gives the propensity to feel the emotion of guilt upon cheating another as an example of a commitment device. The individual who has such a propensity seeks to avoid it and thus avoids cheating others even if they could get away with it. If they can signal such a propensity, others who detect the signal can feel safer in interacting/cooperating with such an individual, thus providing both with benefits. However, Joyce argues that (a) it is not clear that the emotion of guilt *has* associated body language with which to act as a signal, and (b) that guilt is not only an emotion, it requires the belief that one has transgressed a norm and deserves punishment. Joyce suggests we can and do signal moral commitments not just through body language but through action and actual language. Just by acting according to the commitment, due to its costly nature, we signal to others that we are so committed. Furthermore, by making public declarations of our moral judgments and commitments, including making moral judgments of others, we signal our commitments (Joyce, 2006). We can also conceive of such public declarations as somewhat costly, by making a moral judgment, even regarding another person, we indicate that we would deserve punishment if we ourselves were to transgress. If such a transgression were to occur, we would find it more difficult to justify our actions to others and avoid punishment (Joyce, 2006)[18]. Thus, making such public declarations can affect the incentives behind our choices, making transgression less desirable, closing off certain future options.

[18] There has been some recent empirical evidence that people judge hypocrites who condemn immoral behaviour that they engage in more harshly because they falsely signal moral behaviour in a way that is more convincing than simply stating that one behaves morally (Jordan, Sommers, Bloom & Rand, 2017).

We can see in this last example of a publicly declared moral judgment the final benefit of morality, that it links inter- and intra-personal commitments, the public and the private, self-directed judgments and other-directed judgments (Joyce, 2006). When we make a moral judgment about another we generally consider such a pronouncement as binding on oneself. When we make a moral judgement about oneself, we are making a judgment about how we could justify our actions to others. Joyce suggests that morality links such judgments in a way that ordinary emotion cannot.

> No matter how much I dislike something, this inclination alone is not relevant to my judgments concerning others pursuing that thing: "I won't pursue X because I don't like X" makes perfect sense, but "You won't pursue X because I don't like X" makes little sense. By comparison, the assertion of "The pursuit of X is morally wrong" demands both my avoidance of X and yours. (Joyce, 2006; p. 117).

In addition, Joyce (2006) believes that only moral judgements can *license* punishment. They also motivate community members to punish transgressors, by making punishment a putative consideration that cannot be ignored, while motivating transgressors to submit to punishment, through the emotion of guilt. This allows moral commitments to serve as a better mechanism for regulating the behaviour of the community than non-moralised emotions such as anger. Morality's benefit therefore lies in its ability to act as a "social glue" (Joyce, 2006; p. 117). It bonds people together in a shared justificatory framework within which both one's own actions and the actions of others can be evaluated, aids collective decision making and negotiation, and helps to solve common group coordination problems (Joyce, 2006, p. 117).

Stanford (2018) argues that it is this linking together of self-directed and other-directed moral judgments that is the main evolutionary benefit of the externalisation or objectification of morality (the perception of moral considerations as objectively binding features of the world) , the other more focused benefits could be achieved through subjective preferences regarding others' and our own behaviour. He argues that externalised morality acts as a mechanism for correlated interaction.

> It is then proposed that such externalization facilitated a broader shift to a vastly more cooperative form of social life by establishing and maintaining a connection between the extent to which an agent is herself motivated by a given moral norm and the extent to which she uses conformity to that same norm as a criterion in evaluating candidate partners in

> social interaction generally. This connection ensures the correlated interaction necessary to protect those prepared to adopt increasingly cooperative, altruistic, and other prosocial norms of interaction from exploitation, especially as such norms were applied in novel ways and/or to novel circumstances and as the rapid establishment of new norms allowed us to reap still greater rewards from hypercooperation (Stanford, 2018; p. 2).

The important point that Stanford is making is that by adopting an objectified morality, individuals are motivated to monitor the moral views of others and to only interact with those who have similar views, thus protecting themselves from exploitation. On his view, moral judgments indeed aid and support inter- and intrapersonal commitments, allowing individuals to gain benefits from cooperation, but objectified morality motivates individuals to automatically link their self and other regarding judgments and thus avoid exploitation in a world where not everyone will act cooperatively and where the social and normative environment is constantly changing. I will discuss Stanford's view of the role of objectified morality in more detail in section two of this chapter.

Joyce (2005, 2006) argued that morality's main evolutionarily advantageous benefit lay in its ability to act as a bulwark against weakness of the will, encouraging cooperative and prudential action even in situations where it would *appear* that one could gain more by acting uncooperatively. Joyce (2006) further argued that morality provided cooperative benefits by supporting interpersonal commitments and linking self-regarding and other-regarding moral judgments as well. Having established that morality provides certain benefits, benefits which would be threatened upon adoption of a moral error theory[19], Joyce then turns to the question of whether moral fictionalism can help us to retain these benefits, arguing that it can, at least to some degree.

2.1.2. The Efficacy of Moral Fictionalism

Joyce (2005) argues for the conclusion that moral fictionalism can fulfil the evolved function of morality, at least to some degree, on the basis that he believes that engaging with fiction can affect our desires and motivation by affecting our emotional states. He argues that engaging with fictional narratives, whether they are movies, books, oral storytelling or simply daydreaming can produce real emotions, and points out the fact that advertisers often use fictional stories and situations to invoke an emotional connection to their product in order to influence viewers'

[19] Whether agnostic or nihilist.

behaviour. Furthermore, Joyce argues that there are often occasions where engaging with a fiction helps to combat weakness of the will. The example he uses is one where an individual decides to get into shape. Suppose that one need only do approximately fifty sit-ups most days in order to achieve fitness. Joyce argues that on the basis of the truth of this statement, the individual ought to believe it. However, paying attention to this belief endangers his goal; because he need only exercise *most* days, he is constantly tempted to give himself permission to take the day off, perhaps assuring himself that he will make it up to himself another day. This threatens to put him under the necessary number of days or sit-ups needed (Joyce, 2005). So Joyce argues that a better strategy might be to follow a stricter rule such as "I must do fifty sit-ups every day, no more, no less in order to achieve fitness". To sincerely hold such a belief would be to hold a false belief, but Joyce argues that in order to receive the benefit one need not truly believe the proposition, one must simply abide by it. One might rehearse the thought in order to fend off weakness of will, and yet when placed in a critical context and asked whether fifty sit-ups every day are really necessary, one may easily deny the proposition and confirm the truth, that only *approximately* fifty sit-ups *most* days is necessary. Joyce therefore concludes that it is obvious that engagement with fiction, even in moral situations, can help motivate and protect against weakness of will.

Joyce (2005) argues that moral fictionalism is not something that one adopts when one is in the throes of temptation, rather it should be thought of as a 'precommitment'[20]. One does not go into a shop, be tempted to shoplift and then make all the calculations and adopt fictionalism so that one does not (Joyce, 2005). It is obvious that such a method is no better at acting against weakness of the will than simply making calculations based upon what is in one's own self-interest, a method, as noted earlier, likely doomed to failure due to our imperfect reasoning ability. Rather, Joyce suggests that what goes through the moral fictionalist's mind is exactly the same as that which goes through the non-fictionalist. For example, when tempted to steal, they may have the thought "*But stealing is wrong!"* and the accompanying emotional reaction, perhaps disgust at themselves for even being tempted.

The idea here is that the fictionalist has previously committed themselves to using moral discourse, both cognitively and emotively, much in the same way that a non-fictionalist does, but

[20] A 'precommitment' can be defined as the set of thought patterns and psychological mechanisms, instilled in us whether through genetics, our upbringing or both, that push us towards certain kinds of action (in this case, moral action), thus serving in a similar capacity to a commitment device (Frank, 1988). In effect, to hold a 'precommitment' to some set of actions, is to be 'previously committed' to performing those kinds of actions, it is not that one actively decides to commit to those actions in the moment, or that one decides to commit to perform such actions in the future, but, by virtue of one's psychology, one has certain future options cut off (or at least made much less appealing).

also has the disposition to deny that anything, including stealing, is really morally wrong when placed in their most critical context. However, Joyce (2005) emphasises that the idea of someone making a conscious choice to precommit to the moral fiction is an artificial idealisation. Instead, he suggests that it is most likely that the fictionalist would be brought up to think in moral terms, that the precommitment to moral thinking would be put in place by the parents (I would add to this the possibility that the precommitment is at least partly innate). In fact, in a fictionalist society it would not be unreasonable for parents to engage in 'white lies' and encourage moral beliefs in children. Later, when the individual develops greater critical thinking skills and develops a greater understanding, Joyce argues they may come to see moral beliefs as unjustified (perhaps through EDAs) and become error theorists. However, even in such a case, due to how the individual is raised, "these patterns of thought might be now so deeply embedded that in everyday life she carries on employing them- she finds it convenient and effective to do so, and finds that dropping them leaves her feeling vulnerable to temptations which, if pursued, she judges likely to lead to regret" (Joyce, 2005, p. 307). Due to the regulative benefit in carrying on in the same manner as she did before, she sees no reason to adopt moral abolitionism, but she has no reason to cease being a moral error theorist either, and thus turns out to be a moral fictionalist.

Frank (1987) shows that it can be a great advantage to evolve a taste for acting cooperatively when one's partner appears cooperative, exactly because having such tastes solves the 'precommitment problem' where, because neither interaction partner can be assured that the other will be cooperative, they both choose to act uncooperatively to avoid being exploited. However, Joyce (2006) argues that Frank's model treats the conscience "seemingly just as a set of communicable motivation-engaging feelings in favour of and against certain courses of action" (p. 121). Joyce argues that such raw aversions do not suffice for morality, there must be a cognitive component as well (for example guilt requires the belief that one has transgressed against a norm and deserves punishment for doing so). Thus, Joyce considers the necessary precommitment not just as emotional attractions and aversions to certain courses of action, but as a commitment to making moral judgments, where this *includes* both such emotional attitudes as well as certain kinds of thought patterns. Signalling that one has a preference for using and abiding by moral judgments (through emotional displays, public declarations, costly compliance etc.), signals that one has a precommitment to acting cooperatively, allowing others to be sure that they can cooperate safely as well. Therefore, for the moral fictionalist to be successful, they must continue to utilise the thought patterns and emotions that point them towards using and abiding by moral discourse, whilst downgrading their epistemic attitudes towards the cognitions from belief to 'make-belief'. When

they feel guilt, for example, they 'make believe' that they deserve punishment for transgressing the norm.

2.1.3 Fictionalism, Reasons and Internalism

Joyce (2006) argues that any moral naturalist theory needs to account for the practical clout of morality in order to count as a theory about *morality* at all. Part of this practical clout is the idea that moral judgments or propositions are 'authoritative' or 'internalist'. He claims that this is necessary because morality evolved to fulfil a specific function, that of regulating our behaviour and promoting cooperation. Thus, any revised morality lacking practical clout must produce a moral discourse that can meet this function. According to Joyce, a merely reliable contingent relationship between the prescriptions of morality and people's reasons for action is not enough, for if our moral prescriptions are only contingently connected with reasons for action, then it may be the case where having the desire/reason to do something even though it is wrong may mean that one ought to do it. If everyone is aware that the prescriptions of morality and the reasons for action are only contingently connected, then the regulative benefits of the moral discourse would be undermined. On a merely reliable relationship, the temptation to cheat may motivate someone to reduce their desire to act morally so as to give themselves reason to cheat rather than act cooperatively. Thus, Joyce concludes that a moral discourse revised on externalist lines would not be able to satisfy the function of the original discourse[21]. This would give us reason to think that no externalist naturalist theory would be 'good enough' as a theory of morality. Since no externalist naturalist theory is 'good enough', and no internalist naturalist theory has successfully located practical clout in naturalistic facts, and is unlikely to ever do so, Joyce concludes that epistemological challenge is not met and most likely never will be, so belief in moral facts is and will remain unjustified.

However, despite his objections to the moral naturalist, Joyce (2005) makes it clear in his argument for moral fictionalism that moral propositions in fact do not need to provide desire-independent reasons in order for morality to achieve its evolved function; one merely needs a precommitment to using moral discourse and acting accordingly. Joyce believes that we can still receive the benefits of morality, i.e. the benefits of acting cooperatively, while refusing to believe in moral facts, by adopting moral fictionalism. That is to say, we can act as though moral facts exist and provide reasons for action regardless of our desires in ordinary contexts while believing that belief in

[21] Although Joyce (2006, 2016c) notes that this is ultimately an empirical matter that may never be properly established.

such facts is epistemically unjustified in our most critical contexts, hence treating morality as a 'useful fiction'. As mentioned earlier, by adopting a precommitment to acting morally and speaking in moral terms (instilled in us by our parents and/or our genetics) Joyce believes that we can achieve this bulwark against the weakness of the will and thus obtain the regulative benefit associated with morality, all while avoiding the maintenance of false beliefs. So the question is: why cannot the moral naturalist do something similar and claim that a precommitment to moral discourse is enough to ensure it fulfils the same role as the original discourse?

Joyce's (2006) main argument for the moral naturalist theories not being 'good enough' unless they meet the requirement of practical clout, unless they are internalist, is that otherwise the ensuing moral discourse would not be able to play the same functional role. But, the moral fictionalist society, according to Joyce, would not even believe in moral facts (and thus cannot believe that they give desire-independent reasons for action) and yet can[22] achieve the regulative benefit through the adoption of a precommitment (effectively make-believing that the 'moral' rules expressed by the fiction provide desire-independent reasons for action). If this is indeed possible, then it would suggest that a revised moral discourse need not be internalist to achieve its function,; precommitments are also likely effective. Therefore, if Joyce wants to argue that a precommitment is enough to allow moral fictionalism to provide the benefits of cooperation then he must admit that the same precommitment is enough to allow externalist moral naturalism to also provide the benefits of cooperation.

Husi (2014) also argues that if fictionalism can capture the benefits of morality, so too can revisionism. That is, we can achieve the benefits of the practice of morality, which he suggests lies in its ability to realise a 'broad variety of shared interests, projects and ends', through revising the original practice, omitting its errors, and essentially switching from morality to 'shmorality'. Husi agrees with Joyce (2005) that the original moral discourse is reasons internalist (saying that it incorporates inflationary truth conditions or categorical reasons), but denies that it needs to be in order to fulfil its function, and thus rejects the idea that fictionalism is the only viable alternative. This is important, because he suggests that the problems with the fictionalist scheme are 'myriad and well documented'. Even though he does not go into detail about such problems he does leave us with a compelling assumption:

[22] Again, Joyce (2005) argues that it is ultimately an empirical matter as to whether moral fictionalism can be successful in the ways discussed. However, Joyce's argument is an attempt to show why this is not impossible, or even unlikely.

> …the comparatively weak assumption my argument relies upon is that everything being equal, practices proceeding on the basis of truthful attitudes are preferable to ones proceeding on the basis of commonly known false attitudes. Given this, we better stop *make-believing* in what does little work and start *believing* in what does most of the work. (Husi, 2014; p. 88)

If this assumption is correct, and it at least intuitively seems to be so, then we should limit the amount of make-believing we have to do as much as possible. If we have a moral naturalist theory that can satisfy the Good-Reason constraint, and can achieve the benefits of morality without becoming fictionalists and without other substantial costs, then it seems we ought to do so. As mentioned previously, Joyce's (2005) own arguments for fictionalism seem to suggest that this should be possible, although ultimately the answer can only be determined on empirical grounds.

2.1.4 Fictionalism and Revisionary Moral Naturalism

The second point to be made about Joyce's (2005) moral fictionalism is that, like Husi (2014) argues, fictionalism needs to be highly disciplined and focused on what really matters to us in order to procure real benefit, and given this, it ends up looking superfluous. The question is why bother with the fiction at all, why not revise our moral beliefs to focus on what matters and what yields real benefits? According to Husi, fictionalists need to be constantly aware of exactly what the fiction is for, that of securing that which is of value to us, and what is only fictitious, and using that awareness to adjust the fiction and keep it from going off the rails in ways that collective practices are historically wont to do. He points to the common tendency to embroider "even the most mundane of reports" (Husi, 2014, p. 89), arguing that there is "always plenty of internal and fiction-specific pressure in the direction of certain modifications and narrative embellishments that produce better and more enthralling fictions … yet which present a considerable risk of diminishing the fiction's capacity to serve our ends" (Husi, 2014, p. 89). As such, there is a real concern about the stability of fictions, a concern about the tendency for fictions to change in content over time, one that seemingly can only be assuaged by applying strict standards aimed at securing what we really value. Further, which particular moral theory to be adopted as a fiction would be determined based on their practical benefits in regards to achieving what we value. Thus, Husi argues that the more highly focused and disciplined such fictional practices become, the less work the fiction seems to be doing and the more work the awareness of what we value seems to be doing, the fiction itself becomes

basically superfluous. Once we realise this, the door opens to a revisionary approach that revises morality according to this awareness of what we value, cutting away the defective and/or erroneous elements of the moral discourse that motivated us to pursue an error theory in the first place, while keeping the elements that motivated us to pursue the, seemingly unnecessary, fictionalist strategy.

One example of revisionist (and moral naturalist) theory of morality is that of Sterelny and Fraser (2016) which identifies moral facts with facts about cooperation. Sterelny and Fraser essentially agree with Joyce (2006) that morality evolved to regulate cooperative behaviour in humans, although their view is an externalist one rather than internalist and they emphasise the role moral concepts play in tracking facts about human cooperation and the social practices that support it. While they suggest morality's main role is the tracking and enforcement of cooperative facts, they acknowledge that the tracking of facts about cooperation was not the only function that folk morality evolved to fulfil; i.e. folk morality is a 'complex mosaic'[23]. Folk moral concepts that fulfil those functions without fulfilling the tracking function remain unvindicated; many of these functions were not adaptive because they counterfactually tracked the truth about the environment, some for example served to signal group identity. Sometimes, other functions will conflict with truth-tracking and cooperation, pressuring individuals to conform to norms even if they eliminate or erode cooperation (Steleny and Fraser, 2016; p. 21). Hence, they suggest that their account is only a partial vindication of morality; belief in many folk moral concepts turns out to be unjustified. They therefore suggest revising our moral beliefs to only include those that are, namely those that track facts about cooperation. In doing so, we may be able to more effectively track facts about cooperation and thus more easily maximise the benefits from cooperation.

Whilst Joyce (2006) believes that all moral beliefs are ultimately unjustified, he puts forward moral fictionalism as a way to capture the benefits that morality evolved to provide, namely those from cooperation, by relying on our precommitment to folk moral discourse. However, if Sterelny and Fraser (2016) are right that folk morality involves not only concepts that evolved to track facts about cooperation but also concepts that evolved to fulfil other oftentimes non-adaptive functions, then it is possible that a moral fictionalist could modify the 'morality' that he follows to capture a greater benefit than otherwise. In fact, this is what Husi (2014) argues a highly focused and disciplined fictionalist practice would do. In folk morality there are many different concepts, some of them will be adaptive in that they aid the tracking of facts about cooperation, some will actively

[23] "…moral judgments function to signal, to bond, and to shape, not just to track; vindication is only in question with respect to tracking… [and] tracking is only partially successful; moreover, its success may well have varied across time and circumstance." (Sterelny and Fraser, 2016; p. 16).

work against that function (and thus be maladaptive), and some will be neither adaptive nor maladaptive[24]. The moral fictionalist has the aim of adopting a set of rules, a useful fiction, which will grant him access to the cooperative benefits associated with morality. The fictionalist is faced with a choice, to adopt folk morality as his fiction wholesale or to adapt it in order to better meet his aim.

If the fictionalist just adopts folk morality wholesale (by accepting whatever he is already precommitted to for example), then he gets all three types of concept, whether they be adaptive, maladaptive or neither. Because the fictionalist adopts even those platitudes that are maladaptive, the fictionalist fails to capture as great a cooperative benefit as possible. Not only that, but because he is merely accepting blithely whatever folk morality he is precommitted to, but does not truly believe it in his most critical contexts, he lacks the ability to critically assess his currently held 'beliefs', and thus lacks the ability to adapt, grow or refine his set of rules in response to new information and contexts, relying only on what the current folk morality is and what he has been taught. Husi (2014) argues that such a fiction, like many collective practices, is even liable to develop a life of its own, possibly straying far from its original purpose and putting the fiction's capacity to serve our ends at risk. The flexibility of fiction conceivably allows an individual or group to make-believe *anything* as part of the moral fiction, even if it is actually detrimental to the goals of the fiction. Therefore, simply accepting folk morality wholesale while adjusting his epistemic stance from belief to make-belief is not the best the fictionalist could do.

In order to acquire greater benefits of cooperation, the fictionalist may instead choose to adapt his fiction, critically assessing the 'moral' concepts in terms of how adaptive they are in tracking facts about cooperation. By removing maladaptive concepts (and perhaps even those concepts that are neither adaptive nor maladaptive), whilst refining those that are adaptive, the fictionalist can achieve a greater ability to track what behaviours promote cooperation and thus achieve greater benefits. This is not a binary choice of course, a society[25] could choose somewhere else on the spectrum, or choose some other criterion to judge which concepts to include in their

[24] Another way to divide moral concepts could be according to how well they promote one's own self-interest. However, Joyce (2006) makes it clear that our moral discourse fulfils its function to provide us with benefits from cooperation, somewhat paradoxically, by not aiming directly at self-interest. The dynamics of Frank's (1988) model in particular show how a commitment to seemingly imprudent actions can be more beneficial than commitments to always promote self-interest.

[25] One may wonder why we are discussing what a *society* should do, and not what individuals should do. The choice of becoming fictionalist or not, revising our concept of morality or not, is meant to be a collective decision, given that morality provides benefits to the collective. "By asking what *we* ought to do I am asking how a *group* of persons, who share a variety of broad interests, projects, ends – and who have come to the realization that morality is a bankrupt theory – might best carry on" (Joyce, 2005; p. 288)

fiction, but the point is that whatever it is that we think morality helps us achieve, the fiction needs to be guided by such pragmatic concerns in order for it to be useful. It cannot be that, as Husi (2014) puts it, anything goes. For example, it cannot be permissible to assign significance to pain only on certain days of the week, or to pain suffered by a certain gender. Doing so would not suit our collective ends, the satisfaction of which being our original motivation for adopting the fiction in the first place (Husi, 2014). Husi therefore argues that the choice of moral theory that the fictionalist aims to make a fiction out of must be guided by pragmatic concerns; they are unable to point to a particular theory as being morally true since nothing literally has that status for the fictionalist, instead they must guide the concepts in their fiction by what helps us realise what we care about/what promotes cooperation.

As has been shown, the fictionalist needs to revise and focus their fiction and keep it disciplined in order to maximise the benefits received from it and prevent it from going astray. The revisionist also adjusts their discourse in a similar way, but since the concepts and norms included just *are* whatever we collectively value or whatever promotes cooperation, it is far less likely to go off-track. Take the previous example of a set of concepts in the original moral discourse. The fictionalist decides which concepts to include on the basis of how adaptive they are for promoting cooperation and assigns to them a fictional status of producing categorical reasons for supporting or abiding them. Husi (2014) suggests that all the revisionist needs to do is assign a different status to the same set of moral concepts the fictionalist chooses. He suggests that the smoothest way to do this is to retain the original moral vocabulary but supply deflationary truth conditions instead of inflationary ones, i.e. adopting moral reasons externalism over moral reasons internalism.

As an example, the original discourse might call 'harming others' *bad*, and assign the action of causing harm to another a moral status that provides desire-independent reasons against conducting that action (Husi, 2014). The fictionalist would retain this assignment as a make-belief, while the revisionist might instead call 'harming others' *booed* and assign it the literally true status of "that which we have resolved to avoid", or perhaps the status of "requiring maxims against in order to maximise cooperative benefits in our society" (Husi uses the term *booed* for philosophical explicitness, but in ordinary contexts revisionists would continue using the term *bad* just with the new definition, in order to utilise our established precommitment to moral discourse). As can be seen, both fictionalist and revisionist approaches appear to use the same set of norms and concepts, just with different epistemic attitudes. As Husi puts it:

> …once we appreciate how focused and disciplined moral fictions must be in order to stand any chance of procuring real benefits, we are already on our way of granting much of what the revisionist needs, namely certain standards of assertability-conditions governing revised moral discourse which are not construed in terms of some direct correspondence to some reality of categorical normative reasons. (Husi, 2014; p. 92)

It is the highly focused and disciplined nature of the fictionalist discourse that is doing most of the work in procuring its benefits, not the fictional nature of the discourse. As such, the revisionist can dispense with the "fictionalist veneer that cannot fool anybody anyway" (Husi, 2014; p. 92) and still acquire the benefits of the discourse by focusing our attention on what was really motivating the fictionalist in the first place, how well the discourse can serve our ends/promote cooperation.

Husi (2014) mostly only considers Joyce (2005); however Joyce (2006) does look at the possibility of revising moral discourse to be an externalist moral naturalist theory, suggesting that morality without practical clout is superfluous to reasoning about desires and values. If Husi is arguing that we should revise our moral discourse to focus on what is really important to us, then Joyce would probably respond with asking 'why not dispense with discussion of morality altogether and just talk about what we collectively value'? If it is just a realisation of what we collectively value, or what promotes cooperation, that is doing the work, then surely we need not talk in moral terms at all, talk of values and desires should be able to play the same role as the original moral discourse. Some philosophers, such as Stanford (2018), argue that merely strong desires should be enough to be a bulwark against weakness of the will, and Frank (1987) shows that even non-moral emotional commitments can be effective commitment devices. There must therefore be another reason why Joyce argues for moral fictionalism over revisionism or abolitionism, particularly why he believes that thinking about moral judgments as giving desire-independent reasons is necessary for morality to provide its benefits. Namely, the benefit that an externalist revisionist theory seems to fail to capture is the ability of morality to link other-directed and self-directed judgments.

2.2. Fictionalism, Internalism and Projectivism

I have made the argument that if the benefits of morality solely stem from acting as a bulwark against weakness of the will, then Joyce's (2005) argument for fictionalism is equally an argument for revisionism; his argument for fictionalism appears to undermine his argument for the

necessity of reasons internalism. However, Joyce (2006) argues that there is another benefit to morality, one that lies in its ability to link self-directed and other-directed judgments. Stanford (2018) argues that it is only this latter benefit that explains why morality evolved to be externalised or objectified, that is why moral discourse evolved to appear to be internalist and we evolved to perceive moral considerations as objectively binding features of the world. He argues that the ability of morality to automatically link self-directed and other-directed moral judgments motivates correlated interaction, ensuring that altruists will automatically seek out other altruists while avoiding and protecting themselves from exploiters. Joyce adds that the mechanism that makes objectification and thus the linking of self-directed and other directed judgments possible is likely moral projectivism, that humans project moral properties onto their environment in similar manner to how they project colour properties for example.

If this view is correct, then the tendency of humans to project morality likely makes up part of the psychological apparatus involved in our precommitment to morality. Therefore, since the externalist explicitly rejects internalism they would be unable to gain the benefit that objectification of morality supplies, namely its ability to link self-directed and other-directed judgements. Furthermore, by rejecting their natural tendency to project moral emotions onto the world, thus rejecting the objectification of morality, the externalist may be sabotaging their precommitment to morality. The moral fictionalist, meanwhile, faces no such problems, they can simply choose to *make-believe* that moral judgments provide desire-independent reasons, relying on their natural tendency to project moral emotions onto the world as a precommitment, and thus receive the benefits of morality.

I will argue though, that the externalist moral naturalist can adopt part of the fictionalist strategy while not accepting all of it. They can insist, while in their most critical contexts, that moral facts do exist (they supervene on natural facts) but do not provide desire-independent reasons for action, whilst make-believing that moral facts provide such desire-independent reasons in ordinary contexts. I call this variant of the externalist, the fictional-internalist externalist, or FI-externalist for short. By relying on their natural tendency to project morality to supplement their precommitment to objectified moral discourse, the FI-externalist can acquire the benefits of morality that are inaccessible to the ordinary externalist moral naturalist. Furthermore, the FI-externalist has access to all the benefits the ordinary fictionalist can acquire whilst limiting the barely-stable fiction as much as possible.

2.2.1. Moral Projectivism

Earlier, we discussed the reasons that Joyce (2006) gives as to why morality might have evolved, that is what Joyce sees as the adaptive value of morality. But Joyce also attempts to give an account of *how* morality might have evolved, what natural selection did to enable moral judgment. Drawing on a body of empirical evidence, he argues that natural selection manipulated our emotional centres, developing in us a tendency to project moral properties on to the world in a similar way to how we tend to project colour onto the world, a view known as moral projectivism. The minimal version of the view has several aspects as outlined in Joyce (2006; 2009). The first is a claim about phenomenology[26]; that in some sense moral properties appear to be instantiated in the world rather than appear to be our own subjective reactions to the world. The second is a claim about aetiology; that this moral phenomenology is in fact caused largely by emotional activity and not through accurate perception of moral properties in the world. This second claim has a corollary; that such moral appearances are to some extent deceptive i.e. they do not track the truth (Joyce, 2006; pp. 128-129).

Perhaps the best way to consider moral projectivism is with an analogy to projectivism about colour and other sensory modalities. For example, when we see an apple, we do not immediately recognise the apple as producing red sensations in us; rather the redness of the apple appears to be located "in the world". This has an evolutionary function, it orientates us to threats or useful objects in the environment, but in some ways our sensory phenomenology is not always an accurate representation of the world. The colour of an apple and the leaves surrounding it appear far more distinct than the frequencies of light bouncing off the objects would suggest, and this is because of the evolutionary benefits that arise from exaggerating the differences, and thus making it easier to discriminate between the fruit and the surrounding inedible plant matter (Joyce, 2006). In such a case, what is literally true and what is evolutionarily advantageous diverge. Joyce, following Hume, also argues that the mind projects emotions and moral judgments in the same way. He gives the example of pity, arguing that when we see a wounded animal and feel pity, we project that emotion, seeing the animal as *pitiable* and not just something provoking the emotion in us.

> The property of *pitifulness* is, in Hume's words, the "new creation" that your mind has "raised"; it seems as if this a feature of the situation, that your pity is a response to this

[26] Although it should be noted that in Joyce (2009, p. 66), he says explicitly that the use of the terminology 'moral phenomenology', 'seems' or 'appears' does not presuppose any kind of phenomenal character in the sense that philosophers of mind use and intend these phrases.

> property (rather than being actively implicated in its creation), and that someone who looks on indifferently, feeling no pity, is missing something and thus is subject to criticism. (Joyce, 2006; p. 126)

This is not to say that moral projectivism implies that the moral sense is quasi-perceptual. Nothing is actually being projected; it is just a metaphor that attempts to describe our experience of the world as including considerations that demand a certain response from us regardless of our desires or interests (Joyce, 2006). For example, it just seems to us that it is a brute fact that killing babies is morally wrong and condemnable, it does not matter that some people may really enjoy killing babies, such individuals simply appear to be missing something, and appear deserving of condemnation. Moral projectivism therefore has both cognitive and emotional components, while a suffering animal may invoke an emotion of pity which is then projected onto the animal, projectivism also implies an account of how the world appears to the one doing the projecting. When saying "the animal is pitiful" they are asserting that the property is instantiated in the animal, thereby asserting something about the world (Joyce, 2006).

Joyce (2006) provides a range of evidence for the plausibility and likelihood of moral projectivism being true, although as noted in Joyce (2009) more empirical work needs to be done[27]. In what follows, I will make my argument on the assumption that Joyce is correct and moral projectivism is the best explanation of our moral phenomenology. My goal here is to argue against Joyce's assertion that moral naturalism fails as a response to the EDA against moral realism, so if I can accept all his premises (including moral projectivism) and still reject his conclusion then I will be in good stead. But it is worth considering for a moment what happens if moral projectivism is false, can Joyce avoid my argument by dropping it from his premises?

Moral projectivism is meant to be an explanation of how we evolved to make (objectified) moral judgments, how the phenomenology of morality contributes to the benefits it provides without being entirely accurate. As mentioned earlier, it has two parts, each of which could be wrong; a claim about phenomenology and a claim about aetiology. If the claim about phenomenology is incorrect, that is if it is not the case that we view moral considerations as instantiated in the world and as objectively binding independent of our desires, then Joyce's complaints about externalist moral naturalism disappear. Joyce (2006) argued that externalism fails

[27] Joyce (2009) aims to conduct some of the preliminary theoretical work needed for empirical research, attempting to delineate various types of moral projectivism and clarifying their necessary subtheses (including the phenomenological and aetiological/causal claims mentioned above).

to count as a *moral* theory because it fails to account for our intuitions regarding the desire-independent reason producing nature of morality. But if there are no such intuitions, then there is nothing to account for. So Joyce's argument relies on the phenomenological claim.

If however, the claim about aetiology is incorrect, that is if the phenomenology of moral considerations is not caused by emotional activity, this could mean two things, either our phenomenology is caused by something else equally inaccurate, or it is caused by something accurate. If our phenomenology is accurate, that is if moral properties *are* in fact instantiated in the environment and that is why we perceive them as such, then some form of internalist moral naturalism should be possible. But Joyce (2006) argues that natural properties cannot explain why we should see them as producing desire-independent reasons for action. Furthermore, his favoured response to the EDA is moral fictionalism, which does not accept the existence of moral facts let alone that they produce desire-independent reasons for action. So Joyce is committed to our phenomenology of morality being inaccurate.

Now it is possible that our inaccurate phenomenology is the result of something other than emotional activity, but it is hard to see how it could be motivating then. In any case, either our tendency to view moral considerations as being objectively binding features of the environment is necessary to achieve the full benefits of morality, or it is not. That is to say, either this tendency is a necessary part of our precommitment and tends to be motivating, or it is not and does not. If it is not, then we could have stopped at the end of Section 2.1; both externalist moral naturalism and fictionalism can rely on their precommitment in order to achieve the full benefits of morality. For whatever benefits that can be acquired, externalist moral naturalism would be just as capable of achieving them as moral fictionalism, which has to be highly disciplined but is barely stable. Although, in this case, there is also likely no reason not to simply talk in terms of desires and values, beyond mere simplicity of using 'morality' as a short hand. On the other hand, if our inaccurate phenomenology *is* necessary to achieving the full benefits of morality then the following arguments should suffice to show that we likely need only be fictionalists about one aspect of the moral discourse[28], that moral facts provide desire independent reasons for action. Externalist moral naturalism, and an emotional/psychological commitment, would otherwise be sufficient. Stanford's (2018) arguments will suggest that the objectification of morality is only necessary to achieve the third major benefit of morality, that of linking self-directed and other-directed moral judgments. He is, however, mostly agnostic as to what the exact mechanism behind the moral objectification and

[28] Provided we have an externalist moral naturalist account that is compelling on other grounds of course, satisfying the Good-Reason constraint for example.

the linking together of self-directed and other-directed moral judgments is. Thus, while moral projection provides a good explanation for how this linking occurs, my argument should function for whatever this mechanism is.

Stanford (2018) argues that Joyce (2006) provides two independent lines of explanation for the objectification of morality, why we not only evolved to make moral judgments, but why we evolved to project moral considerations on the world rather than take them to be our own subjective reactions to it. Stanford's aim here is to rebut Joyce's reasons for the evolution of objectified morality in order to suggest his own. The first argument Joyce uses is that projection would motivate us more effectively than otherwise, acting as a better bulwark against weakness of the will than mere subjective desires. But Stanford argues that in many cases subjective states *are* strongly motivating, using the example of pain. It seems obvious that even if it were most advantageous to always do a certain action in a certain situation (which Stanford argues could not be the case, there are other evolutionary concerns and motivational impulses that need to be balanced against) an arbitrarily powerful desire to perform that action would be enough to motivate the individual. As Frank (1987) shows, many subjective emotional states could play the role of a commitment device without the need for extra beliefs about the objective state of the world.

However, pushing against this somewhat, is the claim that part of why objectified morality is important for a bulwark against weakness of the will is that objectified morality provides putative considerations that cannot be ignored. This would mean one simply cannot lower their desire to act morally/cooperatively in response to temptation (Joyce, 2006; p. 206). So while one's motivation to act morally may be outweighed by other considerations or desires, it cannot be ignored completely. In such a case, guilt may then motivate the individual to make reparations or to accept punishment and thus allow them to get back on good footing with their community, while the expectation of feeling guilty may further motivate one not to transgress at all. Since guilt depends on the belief that one has transgressed against a norm (Joyce, 2006; p. 104), this option is not open to the individual who does not perceive moral considerations as objective facts about the world[29]. But this response relies on the possibility that one can lower their desire for one course of action in response to temptation, but cannot reduce or rationalise away their guilt itself, which I am not convinced of. Joyce even admits that it may be possible to have desires that are strong enough to resist being

[29] It may be argued that it is possible to feel guilty when transgressing against an institution-dependent norm that requires a certain 'buy in', if morality were more like etiquette for example. But then one may be motivated to do away with their guilt by 'buying out' and rejecting the institution. Perceiving moral considerations as institution transcendent is therefore suggested to provide a better bulwark against weakness of the will (Joyce, 2006).

undermined in this manner (Joyce, 2006; p. 206). Therefore, it is not clear that an objectified morality as a result of the tendency to project moral properties is necessary for ensuring a bulwark against weakness of the will.

The second line of explanation from Joyce (2006) is that it is phenomenologically simpler to project moral qualities into the world than to represent such judgments as subjective responses while motivating the relevant adaptive behaviours just as effectively (Stanford, 2018). This is similar to how it seems phenomenologically simpler to project sensations such as redness and heat onto the world rather than represent such as subjective (Joyce, 2006). The function of these sensory modalities is to orient us towards the environment and roughly track certain features located therein (food, sources of warmth, danger, etc.), so it is more efficient to perceive sensations as being located *in* the environment rather than located within us with a second mechanism that connects these sensations to the environment. This could be contrasted with pain, which has the function of pointing us towards problems in the body. Pain could be considered somewhat projected, in that we feel it in certain locations of the body, but there is also an awareness that this is simply a subjective response to a problem (Joyce, 2006; pp. 127-128). We clearly understand that other people do not feel it too, and that the pain is not a property of the problem, but a response, and thus we often take pain-killers and leave the problem to resolve itself on its own if it is minor enough. If moral projection is more similar to the projection of colour or heat, then it could be simply "the predictable result of natural selection's tight-fisted efficiency" (Joyce, 2006; p. 128). Stanford (2018) objects to this line of thought as well, arguing that, unlike sensory projection, there are significant evolutionary incentives to perceive moral demands differently to others in order to better exploit them. That is, those who perceive moral demands as objective features of the world would be forced to presume that others' experiences of those demands are identical to their own and thus be open to exploitation by those who view the world differently.

This line of reasoning seems strange to me. It is not impossible for colour-sighted individuals to understand that colour-blind people exist, and while many colour-sighted folks would presume others see the same way if they had never encountered the concept of colour blindness, it would certainly seem strange to suggest that they are *forced* to presume such. Thus, it is strange to argue the same of moral projection, especially given the incentives to perceive differently; if there are such, then there also would be incentives to evolve to understand this fact. But Stanford (2018) makes an important point that projecting moral demands onto the world is not necessarily the most evolutionarily efficient way of perceiving such properties. It is only when individuals who do project

moral demands are able and motivated to track and avoid exploiters that projection may become the most viable strategy. Thus, he argues that simplicity cannot be the *only* reason why the tendency to view moral properties as objective features of the world evolved since there are significant evolutionary incentives to perceive moral demands differently. Instead, there needs to be some mechanism of correlated interaction to ensure that those who perceive morality as objectively binding are able to protect themselves from exploitation.

2.2.2. Projectivism and the Objectification of Morality

It is Stanford's (2018) argument that the tendency to view morality as concerning objective demands present in the world motivates exactly this sort of correlated interaction. His argument therefore helps us to see why the externalist moral naturalist is unable to capture the full benefits of morality while the moral fictionalist is able to. In Section 2.1.1, I discussed how Joyce argued that one of the main benefits of morality was its ability to connect other-directed and self-directed judgments and thus provide correlated interaction between altruists. With the addition of moral projectivism, we can now see why this might be possible.

By projecting our emotions and moral judgments onto the world, we take them to be objective features of our environment that not only appear to demand certain responses of us, regardless of our desires or interests, but also of others, regardless of their desires or interests. An action or outcome that is *desired* becomes *desirable*, actions that provoke *disgust* become *disgusting*, punishment against a transgressor becomes *just* and not simply an action that provokes a certain satisfaction in us, etc. An action that is morally *desirabl*e appears to give us reason to desire that action and pursue it regardless of any of our other desires. Not only that but it appears to us to give others reason to desire or pursue it as well, and when they do not, we feel we have reason to rectify the situation and encourage them to do so. Conversely, an action that is morally *undesirable* appears to give us reason to avoid committing that action and reason to prevent others from doing so as well, perhaps by punishment. So not only do some actions *demand* punishment or praise, but require someone to do the punishing or praise. Our propensity to project our moral emotions in this manner therefore provides the connection between commitments, a move from it is *bad for others to harm* (because it might happen to me), to *it is bad to harm*, to *it is bad for me to harm others*, and vice versa.

Since moral judgments (and moral projection) generally involve a conative or emotional component (Joyce, 2006; p. 109), altruists are automatically motivated to look for and avoid or censure exploiters. This monitoring could happen in a number of ways; Stanford suggests gossip about others in moral situations (both real and fictional) plays a key part, but the emotional displays suggested by Frank (1987), or public declarations/actions and costly signalling (Joyce, 2006; Jordan, Sommers, Bloom & Rand, 2017) are likely also important sources of information about the moral commitments of others. Supporting this line, there is some evidence that individuals do seek out interaction partners who share similar moral views while avoiding those whose views differ significantly from our own (Skitka, et. al. 2005. So it would appear that objectified morality potentially holds a significant advantage over a subjective morality in social coordination.

Another important aspect of objectified morality for Stanford (2018) is that not only does it motivate correlated interaction, but it motivates it in new contexts and in environments where the social norms are constantly changing. One could easily imagine that we evolved to be robustly, subjectively motivated to avoid or engage in a particular behaviour and enforce the same in others. But if the environmental conditions were to change, such specific behaviour may no longer be adaptive. While some particular set of social norms may be adaptive to obey in one context they may no longer be adaptive in another (e.g. living in a river valley vs. a desert). By objectifying moral norms with only loose regard to what the content of the moral norms must be, individuals can easily adopt new norms through cultural evolution rather than having to develop new subjective motivations through biological evolution, and further be automatically motivated to apply such norms to others, thereby avoiding exploitation. The treatment of moral considerations as objective features of the world therefore turns out to be a more flexible method of motivating correlated interaction than subjective desires. If moral projectivism is the correct theory of human moral phenomenology then, it would explain how we evolved to be motivated to correlate our interactions, to seek out good cooperative partners and avoid exploiters. Furthermore, if treating moral considerations as objective features of the world provides these benefits in ways that treating moral considerations as subjective does not, then that could suggest that while the moral fictionalist could retain these benefits, the externalist moral naturalist should be unable, or at least find it much more difficult.

An objection could be raised questioning whether objectifying moral norms in this way is necessary to flexibly adapt to new circumstances; could not ordinary social norm psychology suffice? Consider, for example, most norms of marriage, puberty, fashion etc. These vary between cultures

and social groups, and can change quickly, but they do not seem to be moral or objectified in the manner described above. For example, when traveling to another culture, we may easily adapt to the local norms of etiquette, suggesting that we do not see the rules of etiquette as objectively demanded as moral norms. So if the moral norms need to be flexible according to the circumstances, why is objectification necessary? I will outline a few possible responses here, but will not go into much detail, that would have to be for another paper.

Firstly, one might say that yes, the above social norms are not moral, and thus do not need to be objectified. In line with Sterelny and Fraser (2016) the above norms act as markers of social group membership, but the specific content is not necessarily adaptive in the way that the content of moral norms often is. Whereas the content of the moral norm will be adaptive depending on the circumstances, norms of fashion for example, will often only be adaptive insofar as they mark out whose culture you share and thus who is likely to share similar values to you (Sterelny and Fraser, 2016). Secondly, the altruist needs to be protected from exploitation, so they need to automatically be motivated to apply the prescriptions of the norm to others. This is not so much the case with the above social norm examples, what you wear has no bearing on the utility of what I wear. Lastly, in the example of adapting to the local norms of dinner etiquette in another country, it is worth noting that with the rise of globalisation, there may be a sort of meta-norm, 'do as the locals do' or 'when in Rome…'. In the past it is possible that norms of etiquette were more objectified, but as intercultural interactions increased such norms became relativised. A new norm could have arisen, suggesting adopting, or at least respecting, the customs of the locals when travelling. Norms of fashion (e.g. sumptuary laws) have also been moralised in the past, they have been used as indicators of social class, and to step outside the norms, to dress as a different class, would have been seen as a transgression against the social order (Killerby, 2002). As the circumstances changed, such rigid norms became maladaptive and ultimately abandoned.

The responses I have suggested introduce some points of difference between moral norms and other social norms, however as Stanford (2018) argues, there is a spectrum of perceived objectivity (with matters of scientific fact at one end and matters of taste at the other) and as the example of fashion shows, norms can shift in perceived objectivity, becoming more or less objectified. Obviously more work would need to be done to elucidate the reasons as to why moral norms need to be objectified in this way when social norms do not, but I take it that it is not obvious that this question poses a significant problem for the account suggested here; that moral norms are

objectified with only loose regard to content in order to automatically motivate correlated interaction over a wide, and often varying, range of social and environmental circumstances.

2.2.3. Projection as part of 'Precommitment'

If Joyce (2006) is correct that the ability to project our emotions on our environment in this way is necessary for moral judgements to hold practical clout and fulfil their evolved function, then implicit in his account of moral fictionalism is not only that we are pretending that moral facts are real, but that we are pretending that they are both objectively binding independent of any our desires and are instantiated in our environment. Joyce (2005) argues that for a moral fictionalist to acquire the benefits of moral discourse they need to rely on their 'precommitment' to that moral discourse. If humans evolved to project moral considerations onto their environment then at least part of the psychological apparatus involved in forming the precommitment to morality is in fact such moral projection. Moral projection involves a particular way of thinking about and viewing the world that tends to motivate one to act accordingly, which is why Joyce (2006) and Stanford (2018) argue that perceiving moral considerations as objectively binding features of the world is important for individuals to achieve the full benefits of morality, including correlated interaction. We have seen how the first two benefits of morality (bulwark against weakness of the will and interpersonal commitments) can be achieved through subjective preferences, but if Stanford is right, correlated interaction, or the linking of self and other directed judgments can only be achieved through reliance on a phenomenology of objectified morality. Moral fictionalists who wish to capture not only the first two benefits of morality, but also the third, would need to rely on the phenomenology produced by moral projection as part of their precommitment to moral discourse. Even if they understand that they have no justification for believing in moral facts and thus understand that their natural tendency to view moral considerations as instantiated in the world is merely a result of projection and not an accurate detector, they need to *make-believe* that this tendency is an accurate detector. By pretending their original moral phenomenology is accurate they can remain committed to the moral discourse and thus receive almost all its benefits. The externalist however, neither believes nor make-believes that this natural tendency is accurate, and thus cannot receive the third major benefit of morality, that of linking self-directed and other-directed moral judgments.

We discussed earlier that the externalist moral naturalist can also acquire the first two major benefits of morality without having to rely on a barely stable fictionalist attitude to do so. However,

if the third benefit, that of linking self-directed and other-directed moral judgments, relies on our tendency to perceive moral considerations as objectively binding features of the world, then the externalist will have a much harder time in achieving it than the fictionalist. This is because the externalist believes that moral facts exist but explicitly rejects the idea that they provide reasons for action independent of any of our desires. Thus, the externalist explicitly denies the thesis that our intuitions or perceptions regarding the objective bindingness of morality are accurate. Instead they must accept that if we indeed evolved to have such intuitions, then they are merely projected onto the environment, not the result of an accurate detection mechanism. Since the externalist is not a fictionalist, and is committed only to acting upon what they believe is true, the externalist is unable to make-believe in the intuitions provided by moral projection, and thus is unable to fully rely on their precommitment to morality; they can neither believe nor make-believe in morality as being objectively binding and thus cannot use it to automatically link self-directed and other-directed moral judgments.

A moral discourse that is revised according to an externalist moral naturalist theory will likely therefore not be able to completely fulfil the role played by the original moral discourse. On the assumption that a revision of a concept is only acceptable if the discourse it produces can play the same role as the original discourse, this would suggest that externalist moral naturalist theories are likely not sufficient to be theories about *morality*. However, as we have seen it is likely the revised discourse could play some of the role played by the original moral discourse. And it is important not to be too strict regarding how much a revised discourse must fulfil a functional role. This may bring about a higher-level point of indeterminacy, a question of how much does a revised discourse need to fulfil the function of the original discourse in order to be 'good enough'. In fact, Joyce also raises this point:

> Suppose we have used concept ϕ for ten purposes—*U1, U2*, … *U10* (idealizing horribly here, of course)—and suppose that the best imperfect claimant (call it ϕ*) can be used in, say, eight of those ways. We cannot use ϕ* for *everything* that we used to use ϕ for, but we can use it for *most* things. Well, is that close enough? I feel that at this point we can only reiterate Lewis's question: "Who's to say?" (Joyce, 2016c, p. 94)

Since it appears that the concept of morality on a revisionary externalist moral naturalist approach should be able to be put to some of the uses to which we put the original concept but not all, we may be faced with an ultimately undecidable question of 'is this good/close enough?' One's

answer in such a case may be simply a matter of temperament, where there is no fact of the matter in regards to whether the naturalist or the sceptic is correct.

2.2.4. Recap of Possible Responses to the EDA

Let us recap. We currently have four different responses to Joyce's (2006) EDA under consideration. There is moral abolitionism, moral fictionalism and two general varieties of moral naturalism, internalism and externalism. Adopting moral abolitionism would mean abandoning the very real benefits that morality provides. Internalist moral naturalism faces the problem of trying to locate practical clout in some natural property or set of properties as well as the problem that our intuitions about moral internalism can be subjected to the EDA as well. Moral fictionalism could allow us to capture much of the benefit of morality through relying on a precommitment to it but, all else equal, we would prefer to limit the barely stable fictionalist attitude as much as possible. Finally externalist moral naturalism could allow us to capture the first two major benefits of morality without resorting to 'make-belief' but, due to rejection of our internalist intuitions, would be unable to capture the third benefit of morality, that of linking self-directed and other directed moral judgments. The revised discourse would therefore be unable to completely fulfil the role of the original moral discourse. This may give us some reason to think that this revisionary theory is not 'good enough' to count as a *moral* theory, but it may well just be a matter of temperament. However, there is a possible fifth approach not explored by Joyce. One that is in some ways a mixture of the latter two approaches. This approach would follow the externalist line whilst in one's most critical contexts; accepting that there are moral facts but no desire-independent reasons to act accordingly. However in ordinary, everyday contexts, while this approach would still favour accepting moral facts as real, it would also suggest make-believing in desire-independent reasons for action. Effectively, one could be an externalist moral naturalist but a fictionalist about internalism. I call this *fictional-internalist externalism,* or *FI-externalism* for short.

2.2.5. FI-Externalism

We mentioned earlier that a highly disciplined moral fictionalism starts to look a lot like externalist moral naturalism. The fictionalist, in order to keep the benefits of their fiction, of morality, needs to adopt some principles for the government of their fiction, and once they do so,

they appear to be effectively a moral naturalist. However, we now know that even the disciplined fictionalist differs from the externalist moral naturalist in at least one major respect, they make-believe that we have reason to act morally, independent of any of our desires, and this allows them to rely on their projectivist tendencies to link self-directed and other-directed 'moral' judgments, thus achieving the third major benefit of morality. This may make it seem like the fictionalist strategy is better, but the fictionalist needs to rely on a barely-stable fictionalist attitude, one that we have reason to limit as much as possible. Meanwhile, provided their theory is compelling on other grounds (satisfying the Good-Reason constraint for example), the externalist can achieve the first two major benefits of morality while only having to believe in what is true. The only thing missing from the externalist approach is this third benefit.

Instead of abandoning this benefit though, the externalist could adopt a little of the fictionalist approach. In both their most critical and ordinary contexts they can go on believing that moral facts both exist and supervene on natural facts (for example, that certain cooperative actions are morally good because they promote cooperation). But whereas in their most critical contexts they deny that these facts provide desire-independent reasons for action, in their ordinary contexts they could go along make-believing that they do, relying on their projectivist tendencies to do so. This would allow them to continue to act morally, thus receiving benefits from cooperation, whilst automatically being motivated to avoid exploitation by having a poor view towards those who do not act morally as well, even throughout changing societal conditions and norms. This FI-externalist would therefore be able to achieve all three benefits of morality whilst limiting the fictionalist attitude to only that which is unlikely to change. Social norms and societal conditions vary over time and space, and the content of fictions often tend to stray as they are passed from one person to the next, so a moral fiction has to be highly disciplined to keep up. The fictionalist therefore needs to worry about their fiction going astray and work to prevent it (Husi, 2014). But if the phenomenological claim of moral projectivism is true, then we are in some sense 'hard-wired' to perceive moral considerations as objectively binding. So the FI-externalist whose fiction is limited only to "moral considerations are objectively binding" only has to worry about being too enraptured in the 'practical clout' fiction and coming to believe it is true. Furthermore, unlike the fictionalist, they can rely on the property given by their particular moral naturalist theory as a guide to their moral beliefs in changing societal circumstances.

Joyce's (2006, 2016c) argument against externalist moral naturalism is that the revised moral discourse it produces would fail to fulfil the function of the original moral discourse, i.e. we

would not be able to put moral discourse to the same use as before. If this were the case, then we may have some good reason to think that an externalist moral naturalist revision of morality, which rejects practical clout as a requirement or desiderata for morality, would not be 'good enough' to accept. However, we have seen that externalist naturalist theories, provided they satisfy the Good-Reason constraint, should be able to produce moral discourse that can achieve at least some of the function of the original, though the use of psychological precommitments. The knowledge and understanding of the lack of practical clout accompanying moral judgments, however, prevents the revised externalist discourse from satisfying the entire function. The locus of indeterminacy therefore shifts a level; no longer are we asking whether failing to satisfy practical clout, but meeting other criteria, is 'good enough' for a theory about morality, we are now asking whether the revised discourse satisfies enough of the original function to be 'good enough' for acceptance (Joyce, 2016c). It seems that there is no clear answer to this latter question, and there seems no way to find out. The choice, of naturalistic revision or error theory (and then perhaps moral fictionalism) seems but a matter of temperament.

FI-externalism offers a way to avoid this latter question, or at least shift the balance of temperament closer to revision over error theory. By adopting a little bit of fictionalism about practical clout, the externalist can now achieve all three major benefits of morality. In so far as this is the full functional role of the original discourse, then according to Joyce's (2006) methodology for determining whether a revisionary theory is 'good enough', then a FI-externalist theory should be 'good enough' as long as it satisfies our other desiderata for morality.

Now this is ultimately an empirical matter, and perhaps there are other functions of the original discourse that externalism, even with fiction about practical clout, can never capture. Sterelny and Fraser (2016) argue that folk morality is a complex 'mosaic' of (sometimes contradictory) functions, and argue their externalist naturalist theory (identifying morality with facts about cooperation and the social factors that support it) is only a partial vindication of the concept of morality. An FI-externalist version of that theory then would only be a partial vindication of morality as well. But Sterelny and Fraser raise an important point, often the choice is not between elimination and full vindication of a concept; even in scientific domains we may find some parts of a discourse, theory or belief-formation process useful or truth-tracking, while finding others to be debunked. In such cases, revision of the theory or process accordingly is often preferable to elimination. The same can be said for morality and the complex mosaic of functions it provided. In so far as morality evolved to track and promote cooperation and the social practices that support it, a

FI-externalist theory likely could produce a discourse that can play that same role. According to Joyce's (2006, 2016c) methodology, if morality really is a 'mosaic' of functions, such a FI-externalist theory would be a partial vindication of morality. Given that, according to moral nativist hypothesis used in the EDA, the evolved function of morality *is* to promote cooperation and the practices that support it, an FI-externalist theory identifying or grounding moral facts in facts about cooperation could help meet the epistemological challenge posed by the EDA, so long as such a theory is compelling on other grounds (for example, meeting the Good-Reason constraint).

Chapter 2 Conclusion

In this chapter, I discussed and argued for the effectiveness of moral naturalism, specifically externalist moral naturalism, as a response to the Evolutionary Debunking Argument (EDA) outlined by Joyce (2006). In the first section of the chapter I discussed Joyce's (2005, 2006) argument for moral fictionalism in detail, starting with the three major benefits of morality and how moral fictionalism could allow us to continue to capture them even in the face of a moral error theory. I then made use of this argument to show how morality on an externalist moral naturalist theory can also achieve two of the three major benefits. In the second section of the chapter, I introduced fictional-internalist externalism, or FI-externalism for short, to show how we could capture the third major benefit of morality by adopting an externalist moral naturalist theory along with a highly restricted fiction, the make-belief that moral facts provide desire-independent reasons for action. Hence, I have shown that if by adopting moral fictionalism we can retain the three major benefits of morality, then by revising our moral discourse and adopting an FI-externalist theory we can do the same. We should therefore be able to resist Joyce's EDA by adopting an FI-externalist theory, so long as we have independent reason to adopt a particular naturalist theory.

In the next chapter, I will discuss Street's (2006) EDA, and explore whether moral naturalism and FI-externalism can resist her objections to the moral/value naturalist.

Chapter 3:

Moral Naturalism as a response to Street's EDA

In Chapter 2, we discussed and evaluated moral naturalism and moral fictionalism as responses to Joyce's (2006) EDA, and introduced a new approach, FI-externalism, that took aspects from each in order to show that a moral naturalist approach could be a satisfactory response to Joyce's EDA. However, another influential EDA, the Darwinian Dilemma, introduced by Street (2006), targets not just moral beliefs, but evaluative beliefs as well in order to promote Street's own, anti-realist evaluative view, Humean constructivism. Many theorists have ignored the Darwinian Dilemma's targeting of evaluative realism as a whole in favour of discussing its implications for only the moral domain. While it is understandable they may do this, I believe it to be a mistake, as Street's arguments against naturalism rely on the fact the Darwinian Dilemma targets all of our evaluative beliefs, not just our moral ones. Taking this into account, I will be discussing moral/value naturalism as a response to Street's EDA for the evaluative domain as a whole.

In Chapter 1, I outlined Street's (2006) evaluation of the arguments of the value naturalist and her two main objections to those arguments, the 'one level up' objection (Section 1.6.1) and the 'trivially question-begging' objection (Section 1.6.2). In this chapter, I will explore whether externalist moral naturalism, including FI-externalism, can meet the epistemological challenge of the Darwinian Dilemma, ultimately arguing that although it can defeat a version of the Darwinian Dilemma that targets only morality, it is unable to defeat the version that targets evaluative/normative[30] realism as a whole. It may therefore appear that moral naturalism can (at least partially) vindicate moral realism, but is unable on its own to vindicate evaluative realism.

[30] Recall that these two terms are often used interchangeably in the literature. The term 'evaluative' appears to be used to emphasise that we are talking about what *is valuable*, while the term 'normative' appears to be used to emphasises that we are talking about what *reasons we have for action.* However, for authors such as Street, what is valuable is what we have reason to pursue, so the two terms are interchangeable. For consistency and expediency, I will continue to use 'evaluative' to mean either.

3.1 Moral/Value Naturalism and Street's Original Darwinian Dilemma

As discussed in Chapter 1, Street's (2006) EDA poses a dilemma to the evaluative realist who holds that there are at least some evaluative facts or truths that hold independently of all our evaluative attitudes despite the fact evolutionary forces have had a tremendous influence on the content of human evaluative attitudes. The challenge for the realist is to explain the relation, if any, between these evolutionarily influenced attitudes and the independent evaluative truths. Realists can take one of two approaches, they can either assert that there is a relation, or deny that one exists. Taking either horn of the dilemma leads to an unacceptable conclusion for the realist. Denying that there is a relation leads to an implausible sceptical conclusion that our evaluative judgments are likely mostly off-track. While asserting a relation forces the realist to accept the *tracking account*, which is scientifically inferior to the *adaptive-link account*, wherein the truth of the independent evaluative facts plays no explanatory role in our belief formation process. In actuality, the Darwinian Dilemma targets all evaluative beliefs, but Street argues that her anti-realist Humean Constructivist account can avoid the force of the dilemma as it can accept the adaptive-link account. The Humean Constructivist account accepts a version of the broad tracking account (see section 1.4) where the evaluative attitudes track the truth because they are the *grounds* for their own truth (Berker, 2014).

The realist can take a similar strategy, accepting the *broad* tracking account by grounding, or identifying, the moral facts in the some *third-factor account* (see section 1.1. for details); for example, some set of natural facts (Street, 2006; Berker, 2014). If there is some non-normative third category of facts that grounds the independent evaluative truths, and which can be tracked through normal means, it would be no coincidence that we evolved to track the independent evaluative truths via this third category of facts.

According to Street (2006), the Darwinian Dilemma is meant to leave knowledge in many other domains untouched; that is to say, the tracking account is often the correct explanation in other domains. For example, the tracking account gives a good explanation of the relation between our beliefs about midsized objects in our environment, such as trees, predators, cliffs etc., and the truth of these beliefs. The explanation is that we evolved to hold such beliefs because they were true, and being able to discern that truth, to be able to spot that apple tree, spot that tiger in the bushes, was advantageous. Thus, the realist could take a similar approach if they can reduce evaluative/moral facts to some kind of natural facts i.e. provide a compelling value/moral naturalist

account. They can successfully assert a relation between evolutionary pressures on our evaluative judgments and the natural facts that are identical to the independent evaluative truths. “In particular, the relation is this: in ways roughly analogous to the ways in which we were selected to be able to track, with our non-evaluative judgements, facts about such things as fires, predators, and cliffs, so we were also selected to be able to track, with our evaluative judgements, evaluative facts, which are just identical with such-and-such natural facts.” (Street, 2006; p. 136).

So it would seem that the realist can avoid the implausibly sceptical conclusion of the Darwinian Dilemma if they can take a compelling third factor approach, value naturalism for example. However, this is not the end of the story. Berker (2014) argues that there are two main objections to third factor approaches, including value naturalism, in Street’s work. The first is that these approaches only put off the Darwinian Dilemma to a higher level, such that the Darwinian Dilemma can be run ‘one level up’ (See Section 1.6.1). The second is the claim that third-factor accounts are ‘trivially question-begging’, that they must appeal to substantive moral/evaluative truths in order to explain how we were selected to track those truths, begging the question at hand (See Section 1.6.2). Each of these arguments attempts to show that a compelling third factor account is unavailable, and thus that our moral and evaluative beliefs remain unjustified.

We saw in the previous chapter that adopting FI-externalism may be able to help an externalist naturalist theory meet Joyce’s (2006) demands on a successful moral naturalist response to his EDA. The question I want to examine here is whether moral naturalism is able to successfully meet Street’s (demands) on a successful moral naturalist response to her Darwinian Dilemma. I will therefore examine whether a moral naturalist theory can defeat her ‘one level up’ and ‘trivially question-begging’ objections, as well as Street’s (2008) demands regarding the analytic definition of moral realism. I will start with the ‘trivially question-begging’ objection as it can be resolved more straightforwardly than the ‘one level up’ objection. I will show that a compelling externalist moral naturalist theory can largely avoid the force of these arguments, but only at the expense of failing to vindicate evaluative realism.

3.2. Moral Naturalism and the ‘Trivially Question-Begging’ Objection

As mentioned in section 1.6.2, the ‘trivially question-begging’ objection is really a requirement for having a moral naturalist account that meets the *Good-Reason* constraint, rather

than a merely possible one. Essentially, the objection argues that moral naturalists cannot simply assume the truth of substantive moral and evaluative theory, such as a natural-moral identity, in trying to explain how we evolved to track the truth. Instead they must give some good epistemic reasons for believing that the moral facts are in fact identical to or grounded by the natural facts suggested by the theory. Importantly, this objection is only specific to domains of knowledge not targeted by the EDA, i.e. domains of knowledge in which we are entitled to be epistemically conservative about are immune to this objection. It is also important to note, that Street's (2006) anti-realist theory is also subject to this objection and therefore also must meet its challenge as a necessary, but not sufficient, condition to succeed at meeting the epistemological challenge of her EDA (Berker, 2014).

In Chapter 2 of this thesis, I was interested primarily in Joyce's (2006) argument that a compelling moral naturalist account is impossible, or at least extremely unlikely, because no such account can appropriately satisfy our criteria for a theory to be about *morality*. To this end, I was not so much concerned with Joyce's other criteria for a compelling moral naturalist account, instead simply assuming that such criteria could, at least in principle, be met. At least one of these criteria for a compelling moral naturalist account is essentially the challenge of the 'trivially question-begging' objection. That is to provide us with good reason to think the moral facts *are* identical to or grounded in some set of natural facts that we evolved to track, and to provide us with an account of what these natural facts are, rather than a merely possible account of how the moral facts *could* be identical to some set of natural facts that we *could* have evolved to track. Ultimately, it is an empirical question as to whether this *Good-Reason* constraint has been or could be met. However, I will summarise here a moral naturalist theory that seems like it may meet this criterion.

Sterelny and Fraser (2016) suggest that folk moral concepts evolved, in part, to track facts about human cooperation and the social practices that support it. They therefore present a moral naturalist theory where moral truths are natural truths about such facts. That is "moral truths specify maxims that are members of near-optimal normative packages- sets of norms that if adopted, would help generate high levels of appropriately distributed, and hence stable, cooperation profits" (Sterelny and Fraser, 2016; p. 5). However, it only attempts a *partial* evolutionary vindication of the 'folk' conception of morality as identical to facts about cooperation. This is because, according to Sterelny and Fraser, folk morality is a 'mosaic' of different functions, platitudes, beliefs and norms, many of which will end up subject to the EDA, and only some, those which provide an adaptive link between our environment and our behaviour due to aiding the

tracking of facts about cooperation, will be immune. Folk moral concepts that fulfil those functions without fulfilling the tracking function remain unvindicated, and thus belief in such concepts is unjustified. Sterelny and Fraser argue that while their theory may not vindicate all or even most of the platitudes commonly thought to be relevant or even necessary to the folk concept of morality, it is not necessary that it should do so.

Sterelny and Fraser (2016) draw an analogy to the progression of the discipline of astronomy to illustrate their point. While most of the general beliefs involved in ancient Mediterranean astronomical thought were false, agents were nevertheless able to use astronomical info adaptively, e.g. for navigation and time-telling. These astronomical beliefs counterfactually tracked some structural and dynamic features of the solar system quite accurately – sky watchers had a complex of discriminative capacities as well as a complex of explicit (albeit false) beliefs. Thus, in virtue of its ability to provide relatively accurate 'know-how', knowledge of facts about navigation, relative terrestrial locations etc., ancient astronomy acted as a 'fuel-for-success', providing an adaptive link between environment and behaviour. So while many ancient astronomical beliefs were false and remain unvindicated, many tracked facts about the world, and thus could be considered vindicated, leading to an overall partial vindication of the ancient discipline. Over the years, as the discipline was revised, unvindicated beliefs were cast aside, while vindicated beliefs were kept, and new knowledge added, leading to the discipline of astronomy as it is today. The point of this example, as expressed by Sterelny and Fraser, is to show that it is a mistake to frame the question of folk frameworks as one of either reduction or elimination; many if not most cases would involve a mixture of vindication and rejection. Some aspects of the framework are useful in understanding, conceptualising, tracking and navigating the world, whereas others are less adaptive. The 'know-how' is often vindicated whereas the false explicit beliefs could be safely eliminated and replaced.

Sterelny and Fraser (2016) argue that the case for morality is similar to that of astronomy, moral cognition also involves some amount of 'know-how', through the ability to represent and navigate the social environment. Moral judgments made by oneself and others impart information about the social environment, about people's preferences and expectations, about solutions to social coordination problems, etc., and offer a path to successfully navigate this environment, through prescriptions of behaviour, altogether providing a 'fuel for success'. However, only some moral beliefs promote cooperation because they counterfactually track the truth (Sterelny and Fraser, 2016). The truth of these sorts of moral beliefs are conditional on the truth of certain facts about the world, the claim 'it is wrong to murder' is only true if the maxim 'don't murder' is a

member of the 'near optimal normative package' for the society. This can be contrasted with other norms that do not have this property, for example a norm that one must wear a certain hat only promotes cooperation in that it acts as a marker of group identity; the content of the maxim has no effect on cooperation. One could replace the hat with any other type of hat, or even something else entirely, and as long as everyone else did the same, the effect on cooperation would be the same. For the moral claims that are true, it is the content that makes them so; if it were not the case that the prohibition against murder was part of the 'optimal normative package', then the belief 'it is wrong to murder' would not promote cooperation. The counterfactual sensitivity of these sorts of claims is what makes them appear to meet the epistemological challenge of the EDA. However, only some moral claims hold this kind of counterfactual sensitivity, folk morality also contains the sort of claims that do not. So folk morality remains only partially vindicated. A revision of the discipline therefore may excise those beliefs that are unvindicated, while keeping those that remain vindicated.

I want to make it clear that my endorsement of Sterelny and Fraser's (2016) account is only tentative. The preceding discussion is less an endorsement of the theory and more to provide an example of an account that seems to satisfy at least some of what we want out of a compelling naturalist theory in terms of avoiding the trivially question-begging objection. Sterelny and Fraser's theory attempts to provide an actual account of how the moral facts are identical with a set of natural facts and how and why we evolved to track them, explaining what exactly the natural facts are and how our moral beliefs served to track them. Furthermore, Sterelny and Fraser provide empirical evidence to back up their claims, rather than presenting their account as a mere possibility. The theory therefore provides us with some good reason to believe that it holds. Furthermore, because Sterelny and Fraser admit that their theory only partially vindicates morality, it is clear that they are not just assuming that large swaths of moral theory are true. Instead, their approach examines what sorts of moral content appear counterfactually sensitive to the environment, and then making an argument that it is only those sorts of moral beliefs that are vindicated. The result is a moral naturalist theory that appears to avoid the 'trivially question-begging' objection.

For such a theory, the 'trivially question-begging' objection is no impediment for it (at least partially) vindicating morality and moral realism, restoring justification in at least some of our moral beliefs. Opponents either need to challenge the evidence they provide, or they need to challenge how compelling such a theory is on other grounds. For example, it could be said that Sterelny and

Fraser are still making the assumption that our evaluative beliefs about morality and what functions and concepts it involves, our metaethical beliefs, are roughly reliable. Their theory therefore can still be subjected to the Darwinian Dilemma 'one level up'.

3.3. Moral Naturalism and the 'One Level Up' Objection

In this section I will show that while moral naturalist theories may appear to resist the version of the Darwinian Dilemma that is often discussed (the version that targets only moral realism), they in fact are forced by the 'one level up' objection to face the Darwinian Dilemma that targets evaluative/normative realism as a whole. This is similar to how value naturalist theories, while resisting the original Darwinian Dilemma, must then face the same dilemma 'one level up' (Street, 2006). Even so, I will argue that we can in fact (at least partially) vindicate[31] moral realism by making use of what Street (2006) calls the 'rigidifying move', but only at the cost of failing to vindicate evaluative realism. However, externalist moral naturalism does not even try to vindicate evaluative realism and thus is not targeted by the original Darwinian Dilemma which targets evaluative realism. Therefore, the rigidifying move should be a satisfactory method of resisting the 'one level up' objection. Moral realism can seemingly be vindicated without having to vindicate evaluative realism first.

While moral naturalism aims to vindicate moral realism, according to Street's (2006) 'one level up' objection it must vindicate evaluative realism to do so. The 'one level up' objection stems from Street's insistence that the truth of the grounding relation (G) utilised by third-factor theorists (including naturalists) also counts as an evaluative truth (Berker, 2014). For example, the moral naturalist might argue that the moral facts supervene on some set of natural facts, and thus a satisfactory broad tracking account explanation can be given. But in order to determine this truth, they must rely on their substantive moral theory and/or their evaluative beliefs. To make a judgment of which natural facts ground the moral facts, one needs to know a little something about what morality is, to do that one must rely on their evaluative judgments regarding morality. This means that the truth of the relation (G) (Non-normative Fact F (at least partially) grounds normative fact N) counts as an evaluative truth on Street's account. Since our evaluative beliefs are likely heavily saturated with evolutionary influence, then whatever method is used to determine that (G) is true would also be heavily saturated with evolutionary influence. The Darwinian Dilemma then arises

[31] For ease of expression, I will simply write 'vindicate' to mean '(at least partially) vindicate' for the rest of this section.

with the question, what is the relation between our evaluative attitudes shaped by evolutionary forces and the independent evaluative truth of (G), the truth of the natural-evaluative identity in the case of the naturalist. Street argues that finding an appropriate third-factor account is unlikely at this level, and even if they could, they would once again be subject to the Darwinian Dilemma the next level up. Meanwhile, the narrow tracking account is scientifically inferior to the adaptive link account, which need not posit the stance-independent evaluative truth (Street, 2006; p. 141). FI-externalism would not help here; nothing about adding the fictionalist stance of make-believing in desire independent reasons for moral action while in ordinary contexts makes it any more successful than ordinary externalist moral naturalism in this regard.

Take Sterelny and Fraser's (2016) externalist moral naturalist theory for example. We have already seen that Sterelny and Fraser can provide an externalist moral naturalist account that can satisfy some of our criteria for a compelling naturalist account. Let us assume then that it is successful at meeting the epistemological challenge of the original EDA, giving us good reason to think that at least some of our moral beliefs, namely the ones that help us track facts about cooperation and the practices that support it, are vindicated. However, this account is still an account of morality in terms of its function; it relies on our intuitions and evaluative judgments regarding what the purpose of morality is i.e. promoting cooperation and the practices that support it. The argument that Street (2006) makes, is that this is just one possible option that morality, or the evaluative truth, could be. It is conceivable that what is really valuable is simply what kind of hat we wear, regardless of its effects on cooperation in society. By making a judgment that cooperation is what is important when determining what the moral facts are identical to, we are making an evaluative judgment, a judgment that is influenced by evolutionary forces. So the naturalist attempts to vindicate morality with a story of how moral beliefs evolved to track moral facts by tracking facts about cooperation, but they need to give a story of how they evolved to track the truth of the *evaluative attitudes* used in developing this story, in making the judgment that moral facts supervene on facts about cooperation.

Sterelny and Fraser (2016) might reply that we did not simply *evolve* to make this judgment, given that folk morality is a 'mosaic' containing many judgments with varying purposes, but rather we use scientific methods and rational reflection to come to this conclusion that a segment of our moral beliefs have the purpose of promoting cooperation due to their content. We can 'see' that morality has a function of regulating behaviour in society in order to provide cooperative benefits, and we can see that only some moral beliefs do this because of their content. But why, out of all the

functions contained in the 'mosaic' of folk morality, choose that particular function to revise our conception of morality around? Is it because the claim 'we should revise our conception of morality to only include the function of promoting cooperation and their benefits cooperation provides' is true? Or is the adaptive link account correct, that we value cooperation and the benefits it provides highly, and thus we value this particular function more highly than others, because it provided an adaptive link between the environment and our behaviour, and it would do so regardless of the truth of that claim?

What this amounts to is a Darwinian Dilemma 'one level up'; we are moving from a dilemma targeting moral realism to one targeting evaluative realism. The answer to the question 'what is the relationship, if any, between our evolutionarily influenced moral beliefs and the stance independent moral truths?' is dependent on answering 'what is the relationship, if any, between our evolutionarily influenced evaluative beliefs and the independent evaluative truths?' which is just the original Darwinian Dilemma. This is why it is a mistake to attempt to limit the scope of the Darwinian Dilemma to just moral realism, to answer that dilemma satisfactorily one needs to answer the dilemma targeting evaluative realism satisfactorily. The moral naturalist could therefore attempt to either broaden their account to explain the supervenience of the evaluative on the natural, or they can attempt some other third-factor account of how the evaluative is grounded by the non-evaluative. Whichever path they take though, they will be targeted by the original 'one level up' objection; they need to not only provide a third-factor account, but also to provide a story of how the evaluative attitudes used in determining the truth of that third-factor account are reliable.

FI-externalism does not resolve this issue. Adding the fictionalist aspect to a standard externalist moral theory says nothing about what grounds the evaluative attitudes that push us towards accepting the particular moral naturalist theory or to adopting the fiction of internalism on top of it. It therefore cannot even defend the moral naturalist theory from the original Darwinian Dilemma, let alone the dilemma one level up. So it would seem that in order to answer the Darwinian Dilemma that targets moral realism, one must first answer the dilemma that targets evaluative realism as a whole, and if one can successfully answer that dilemma, then one has already successfully vindicated moral realism as well.

3.3.1. The 'Rigidifying Move'

The move from the dilemma that targets moral realism to the dilemma that targets evaluative realism can be resisted by taking an approach Street (2006) calls 'rigidifying'. Street (2006) actually discusses this move only in regards to the attempt to vindicate evaluative realism, arguing that it is ultimately unsuccessful, failing to ensure a given value naturalist theory is genuinely evaluatively realist on her taxonomy. However, I argue that this approach can be more successful when used to vindicate moral realism, as long as the realist is not committed to moral facts being reason-giving independent of any of our desires.

This is the rigidifying move as Street (2006) puts it:

> Consider, for instance, a view which says that which natural facts evaluative facts are identical with is fixed in some way by our actual evaluative attitudes (in other words, by our attitudes, here and now). And suppose that our actual attitudes determine it that the evaluative facts are identical with natural facts N. On such a view, even if we had had entirely different evaluative attitudes, it still would have been the case that the evaluative facts are identical with natural facts N, since those are the ones picked out by our actual evaluative attitudes. (Street, 2006; p. 138)

The goal of the rigidifying move is to make the natural-evaluative identity into a 'rigid designator', i.e. a definition that applies across all possible worlds, by fixing the referent of what it is to be an evaluative fact to being a member of the set of natural facts, N, i.e. the natural facts that our actual evaluative attitudes suggest are identical to the evaluative facts (Lewis, 1989, p. 132; Street, 2006, p. 30). Effectively, 'evaluative facts' becomes a name that picks out a particular class of objects across possible worlds, regardless of what people in those worlds think the natural facts that the evaluative facts are identical to are. This is similar to how the name 'Aristotle' picks out the same individual in all possible worlds, even in ones where the individual went by a different name (Gendler and Hawthorne, 2002).

An analogy to the naming of other natural kinds may be useful here. For example, what natural kind does the term 'heat'[32] pick out? We might say that 'heat is the phenomenon that generally produces sensations of warmth'. However, 'the phenomenon that generally produces

[32] Example based on one from Gendler and Hawthorne (2002).

sensations of warmth' is not rigid; it picks out different things in different worlds, in our world that would be molecular movement[33], in some other possible world that might be something else. We might therefore want a rigid term that we can use to discuss a particular natural kind no matter the world we are talking about. We might want to ask for example, 'what if heat produced some other sensation, pressure for example?' How would we make sense of that sort of question?

One method is to 'fix the reference' of the name that's intended to be rigid by use of a description that is not. We fix the referent of 'heat' by how it is used in the actual world, what 'the phenomenon that generally produces warmth' is in the actual world, and that is molecular movement. So we could substitute molecular movement into our question and get 'what if molecular movement produced some other sensation, pressure for example?' The question now makes sense. It is likely we do this with most natural kind terms (Gendler and Hawthorne, 2002; p. 29), fixing for example the reference 'light' by the visual appearance it produces, 'sound' by the auditory experience etc. It might be thought that a similar community in another world could do the same thing, fixing 'heat' according to what produces sensations of warmth in their world. However, there would be no contradiction; these two uses of the word 'heat' are referring to two different concepts. If evaluated from a third-party, independent perspective, they could be relabelled to avoid confusion (perhaps 'heat-prime' and 'heat-alpha' for example). But we are not looking at them from a third-party, independent perspective, so which term we should use is determined by which is most useful. Given that we, in our world, use 'heat' to mean 'molecular movement' and not something else, it seems plausible to think that that usage is what we find most useful (hence the reduction in the first place), so that is the term we should use, even if talking about the natural kind in other worlds.

In the case under discussion now, 'evaluative facts' falls into the same role as 'heat' (the rigid designator), 'the natural facts N' holds the same role as 'molecular motion' (the natural kind) and 'the natural facts that our evaluative attitudes pick out as identical to the evaluative facts' has the same role as 'the phenomenon that generally produces sensations of heat' (the non-rigid descriptor). The value naturalist who takes the rigidifying move fixes the referent of 'evaluative facts' according to what satisfies the non-rigid descriptor 'the natural facts that our evaluative

[33] 'Molecular movement' might seem like a simple enough way of referring to the natural kind in question, so the question might be raised 'why not use the term 'molecular movement' instead of 'heat'?' But even 'molecular movement' is an abstraction and a simplified, deliberately fixed reference for a complex phenomenon. In many cases the natural kind in question may be a very complex phenomenon, possibly even a Boydian 'homeostatic cluster' (Boyd, 1988). It is therefore often useful to have a simple way to refer to these complex phenomena to ensure ease of communication.

attitudes pick out as identical to the evaluative facts' in the actual world, i.e. 'the natural facts N'. Thus, the value naturalist uses the term 'evaluative facts' to pick out the set of natural facts, N, even in other worlds, and even if communities in those other worlds use that combination of letters and sounds to pick out some other natural kind.

Street's (2006) problem with the 'rigidifying' move when applied to value naturalist theories is that they fail to count as genuinely realist on her taxonomy. This is because other communities with substantially different evaluative attitudes could also pull the same rigidifying move, identifying the evaluative facts with some other set of natural facts. In such a case there would be no robust sense that this alternative community would be making a mistake or missing something.

> And the upshot is that when we say "The good is identical to N" and they say "The good is identical to M." we will not be disagreeing with each other, with one of us correct and the other incorrect about which natural facts the good is identical to, but rather simply talking past each other, with the reference of our word "good" fixed by our actual evaluative attitudes, and the reference of their word "good" fixed by their actual evaluative attitudes... there is, on such a view, no standard independent of all of our and their evaluative attitudes determining whose sense of the word "good" is right or better… (Street, 2006, p. 138)

However, as we have seen with the example of 'heat', this is not ordinarily a problem for natural kind terms. Although a community in another world may be using the same combination of sounds and letters, 'heat', to refer to some other natural kind to us (perhaps because something else causes 'hot sensations'), it is simply a different concept, and which one is better depends on what we find most useful. A third, independent party might relabel the two terms for clarity. So why cannot we take the same approach here, relabelling the terms 'good-n' and 'good-m' for the benefit of some hypothetical, independent third party?

The issue is that 'the evaluative facts are members of the set of natural facts N' is not the entire definition of 'evaluative facts'. Street (2006) argues that when two communities of genuine realists, even ones where the word 'good' is used differently, disagree as to what is 'good', they are in actual disagreement as to what we have reason to do independent of any of our evaluative attitudes.

> … a genuinely realist version of value naturalism will hold that even if the two communities' uses of the word ''good'' track different natural properties, the communities are nevertheless (at least potentially) using the word ''good'' in the same sense - genuinely disagreeing with one another about the correct natural-normative identity - and that there is a fact of the matter about which (if either) of us is right that obtains independently of all of our and their evaluative attitudes. (Street, 2006; p. 139)

The point is that evaluative facts (aka normative facts) are by their very nature supposed to give us reasons for action, and for the evaluative realist, these reasons are supposed to apply regardless of our evaluative attitudes, even across worlds, such that someone who does not hold evaluative attitudes that match the evaluative facts appears to be missing something. So there is a contradiction involved. By taking the rigidifying move a community essentially asserts that they are not disagreeing with another community who fixes their definition of 'evaluative fact' according to a different set of natural facts, rather they are simply using a different concept. However, to be a realist about evaluative facts one must assert that 'there are reasons for action that are independent of anyone's evaluative attitudes'. So the two communities, in asserting different natural-evaluative identities, do appear to be in disagreement, a disagreement about what reasons for action hold independently of everyone's evaluative attitudes and what grounds those reasons[34]. This contradiction does not appear in cases of non-evaluative natural kinds like 'heat'. Such terms, being non-normative, do not assert reasons for action, let alone reasons for action that apply to everyone regardless of their evaluative attitudes. So when two communities fix the definition of a natural kind term differently, they *are* talking about different concepts, but just using the same combination of letters and sounds to refer to these different concepts.

Furthermore, when making the decision of whether or not to take the 'rigidifying' move and how to fix the referent, the value naturalist is making an evaluative judgment, just as the realist about 'heat' makes an evaluative judgment about how to fix the referent of the term 'heat'. A community of realists who rigidify the natural-evaluative identity according to their own evaluative

[34] Lewis (1989) elucidates a similar worry with the rigidifying move in how it fails to do away with the contingency of valuing:

> The trick of rigidifying seems more to hinder the expression of our worry than to make it go away. It can still be expressed as follows. We might have been disposed to value seasickness and petty sleaze, and yet we might have been no different in how we used the word 'value'. The reference of 'our actual dispositions' would have been fixed on different dispositions, of course, but our way of fixing the reference would have been no different. In one good sense – though not the only sense – we would have meant by 'value' just what we actually do. And it would have been true for us to say 'seasickness and petty sleaze are values'. (Lewis. 1989; p. 132 – 133)

attitudes thereby makes the evaluative judgment 'we should (in a prudential sense) fix the natural-evaluative identity according to our own evaluative attitudes'. Since they have no basis to criticise a similar community in another world doing the same according to their own evaluative attitudes, perhaps the principle should be 'a community should (in a prudential sense) fix the natural-evaluative identity according to their own evaluative attitudes'. If this is the case, then the value naturalist is effectively asserting the evaluative judgment that the natural-evaluative identity for a given world/community is dependent on the evaluative attitudes of that world/community. Since Street (2006) insists that "…in order to count as realist, a version of value naturalism must take the view that facts about natural-evaluative identities (in other words, facts about exactly which natural facts evaluative facts are identical with) are independent of our evaluative attitudes" (p. 137), the value naturalist account that takes the rigidifying move in this way would fail to count as evaluatively realist.

The question is then, is morality on a moral naturalist theory more like evaluative facts or natural kind terms? Ultimately, it would depend on whether the theory in question is internalist or externalist. If the theory is internalist, and moral facts provide reasons for action independent of any of our desires/evaluative attitudes, then taking the rigidifying move puts one in the same situation as the value naturalist who takes the rigidifying move. By taking 'provides reasons for action independent of any of our evaluative attitudes' to be part of the concept of a moral fact, the moral realist runs into problems when they fix the referent of 'moral' to a set of natural facts determined by their actual evaluative attitudes. The moral realist ends up asserting that their use of the term 'moral fact' expresses a different concept than a similar community of realists that rigidifies a different natural-moral identity, yet both are arguing about what reasons people have for action, independent of any of their evaluative attitudes.

This is even more clear if we take Joyce's (2006) requirement of 'practical clout' (aka *inescapable authoritativeness*) as central to the concept of a moral fact for the internalist moral naturalist. If this is the case then moral facts would be inescapable, that is they apply to everyone (even across possible worlds), and they are authoritative, that is they provide reasons for action regardless of anyone's evaluative attitudes. Therefore, two communities of internalist moral naturalists from different worlds with substantially different evaluative attitudes will not only accept two different natural-moral identities, two different definitions of the 'good', but each asserts the existence of a set of reasons for action that apply not only to themselves (independent of their actual evaluative attitudes) but also to the other (independent of *their* actual evaluative attitudes).

This leaves at least two different, most likely conflicting, sets of standards of action that are both meant to be inescapably authoritative. But, by definition, it cannot be the case that there are two sets of inescapably authoritative standards of action, so the rigidifying move for the internalist moral naturalist results in a contradiction.

For externalist moral naturalist theories however, the story is very different. This kind of theory lacks the need to be evaluatively realist, for it does not include 'provides reasons for action independent of our evaluative attitudes' in the concept of a 'moral fact'. The rigidifying move ends up looking much more like other natural kind terms. Just like we might fix the referent of 'heat' according to what produces sensations of heat in us in the actual world, we might do something similar with morality:

(1) Moral facts are whatever generally produces moral emotions in us[35]

In our actual world we might have some theory about what that is. If Sterelny and Fraser (2016) are right, that might be facts about cooperation and the social practices that support it. So fixing the referent accordingly, we end up with

(2) Moral facts are identical to the set of natural facts N (for example, N might be facts about cooperation and the social practices that support it in the actual world)

Now of course a counterpart community in another world might find that moral emotions are caused not by N, but by M. If they also take the rigidifying move, they would not be talking about the same thing, they are picking out some other natural phenomena with their use of the sounds and spelling of 'moral'[36]. A third, independent party might relabel the two terms for clarity ('moral-n' and 'moral-m' maybe). Which one, if any, should be used would depend on the population being considered. In our actual world, since we find associating N facts with morality to be useful (since we can make use of our precommitment etc.) and are unlikely to find M facts useful, we should

[35] Given that morality could be considered a 'mosaic' of both truth-tracking and non-truth-tracking functions (Sterelny and Fraser, 2016), we might instead say that 'Moral facts are whatever is picked up by the vindicated truth-tracking discriminative capacities of our moral faculty'.

[36] If neither community took the rigidifying move, it might be possible to say that the two communities are discussing the same thing on the basis of other characteristics of 'morality'. For example, if each use of the concept of 'morality' serves a similar purpose (for example acting as a bulwark of the weakness of the will), 'moral' judgments in both worlds often produce moral emotions etc., then it may be possible to come up with a unified moral theory that accounts for the difference in content, grounds etc. via taking into account the difference between the two worlds.

continue as we are, using 'moral' to refer to N facts. Here the use of 'moral-n' will not necessarily make demands on what people in other worlds where conditions may be vastly different should do.

Now, an externalist moral naturalist theory that takes this approach is obviously not evaluatively realist in the sense Street (2006) is referring to[37], but it is still morally realist, just as in the example of 'heat', we are realists about heat. Heat really exists, independent of any of our desires, because molecular motion exists independent of our desires. Same for morality, it exists independent of our desires, because the natural facts, N (perhaps facts about cooperation and the practices that support it in our actual world), exist independent of any of our desires. Thus, while the rigidifying move is not open to the internalist moral naturalist, it remains open to the externalist moral naturalist, as they do not need to be committed to moral beliefs being stance independent evaluative beliefs and therefore do not need to be committed to evaluative realism. Consequently, just as our beliefs about heat are not targeted by the Darwinian dilemma, neither would our beliefs about morality on an externalist moral naturalist theory, so long as we have other good reasons to adopt a particular moral naturalist theory. Only internalist moral naturalist theories would be targeted, as they aim to be evaluatively realist as well. This leads us to Street's (2008) argument against externalist moral naturalism, that while such theories may count as realist theories about *something*, they nevertheless fail to count as a realist theories of *morality,* because they fail to be evaluatively realist.

3.4. Moral Naturalism and the Desiderata constraint

As we have seen in the previous section, the 'rigidifying' move can rescue externalist moral naturalism from the 'one level up' objection, at the expense of evaluative realism in the sense targeted by the original Darwinian Dilemma. A compelling externalist moral naturalist theory therefore should be able to meet the epistemological challenge of the Darwinian Dilemma against morality. The question then is whether a compelling externalist naturalist theory is possible, that is can a moral naturalist theory possibly meet the 'trivially question-begging objection' or the Good-Reason constraint, and can a moral naturalist theory that is unable to vindicate evaluative realism actually count as a theory about *morality*, meeting the Desiderata constraint. We discussed the first part of the question in Section 33, coming to the conclusion that it seems likely that the 'trivially question-begging' argument could in fact be met. In this section, I will turn to the latter part of the

[37] Although it may be normatively realist in the sense meant by Copp (2009).

question by arguing that our concept of morality can be at least partially vindicated in the face of the EDA and this is likely sufficient to be able to continue to use moral terms. Furthermore, adopting FI-externalism may help ameliorate the practical costs of adopting a revisionary approach to morality.

As discussed in Chapter 1, Street (2008) can be read as making the case that naturalist theories must count as evaluatively realist in order to be *realist* theories of *morality*. It is worth taking this argument seriously as it is similar to the argument made by Joyce (2006) that practical clout, or *inescapable authoritativeness*, is part of the analytic definition of morality, and thus any naturalist theory must account for practical clout in order to be a theory about *morality*. For the purposes of our discussion, I will assume in this section that this is indeed the argument made in Street (2008), even though, considering her later works (such as Street (2012)), I am not sure that this argument was her intention.

Considering the similarities, it is possible that the argument that appears in Street (2008) can be dealt with in a similar way to how we dealt with the argument made by Joyce (2006) in Chapter 2. Additionally, given that Street (2012) rejects internalist definitions of morality, the arguments she provides may be useful in dealing with the objections to the externalist naturalist made by both Joyce (2006) and Street (2006).

3.4.1. Indeterminacy in the Analytic Definition of Morality

Both Street (2008) and Joyce (2006) take practical clout, or uncompromising normative realism, to be part of the analytic definition of morality, as a key desideratum that any naturalist theory must satisfy in order to be a theory about *morality*. If this is the case, then it would be analytic that externalist naturalist theories, which reject practical clout, would fail to be theories about morality; instead they would be realist theories about some other concept, for example a 'schmorality'. Street (2008) argues that "A version of naturalist realism that fails to [have implications about how we have reason to live] is perhaps realist, but not *normative* realist…" (p. 224)[38] and "the whole point of uncompromising normative realism is that it *vindicates* morality if correct..." (p. 223). Even if we can avoid the Darwinian Dilemma 'one level up' applying to externalist naturalism by fixing the natural-'moral' identity according to our actual evaluative attitudes, we may fail to vindicate morality, actually ending up with a natural-'schmoral' identity instead. The point is, if

[38] It should also be noted that Street (2008) occasionally equivocates between 'normative' realism and 'moral' realism.

Joyce and Street are correct, that part of the rigid designator for 'moral judgments' is that they 'provide reasons for action independent of any of our desires/evaluative attitudes', then any theory that fixes the natural-'moral' identity according to our actual evaluative attitudes would fail that criterion, and thus would end up not being about morality at all.

Of course, externalist moral naturalism denies that practical clout is a necessary component of the analytic definition of morality at all. Furthermore, folk morality is a 'mosaic' of various functions, platitudes and definitions (Sterelny and Fraser, 2016). The definition of the term 'moral' can therefore be considered equivocally analytic, or conceptually vague, exhibiting both semantic variation and indecision, much like how the definition of 'value' can be considered equivocally analytic, or vague (Lewis, 1989). While it may be true that morality with practical clout fits the folk use of the term best, there are many imperfect claimants, many of which satisfy nearly all of our other desiderata for use of the term. In the absence of a theory of morality that can satisfy the analytic definition that best fits our folk morality, as well as meet the challenge of the EDA, the term 'moral' may well go to one of these imperfect claimants, those concepts we might otherwise call 'schmoralities' (Joyce, 2016c). However, it still may be said that strictly speaking there is no morality, or no moral facts realistically construed, so an error theory or an anti-realist theory of morality are also potential options to be considered.

A point of indeterminacy therefore arises regarding how best to respond to the realisation that there is no perfect deserver of the name 'morality'. As mentioned previously, resolving this indeterminacy is no easy task. Lewis (1989) argues that it may well be just a matter of temperament. Those with a more error theoretic bent, such as Joyce, may argue for an error theory, others may lean towards revision, endowing some realisable imperfect claimant with the name 'morality'. Whether an externalist moral naturalist theory can meet the Desiderata constraint, providing a claimant that can satisfactorily serve as 'morality', will therefore depend on whether this indeterminacy can be resolved and in what way. Joyce (2006) and Street (2008) appear to make the case that this indeterminacy can be resolved in the error theoretic's favour.

3.4.2. Street (2008) and the Function of Morality

As discussed in Chapter 2, Joyce's (2006, 2016c) strategy for resolving this indeterminacy was to consider whether the revised discourse can play the same role as the original discourse once

did. If it can indeed fulfil the function of the original discourse then the revisionary approach is likely acceptable. However, if it cannot, then an error theory should be preferable. He argues that in the case of morality, revisions of the discourse according to externalist naturalist theories fail to fulfil the function of the moral discourse. I devoted Chapter 2 to showing that a revision according to a moral naturalist theory that meets the Good-Reason constraint, should, in fact, be able to fulfil the function of the original moral discourse.

What about the argument in Street (2008)? Does it make any claims about the function of moral discourse? In parts, Street (2008) does seem to argue that a moral discourse that does not allow for reasons *simpliciter* for action would seem strange and perhaps impair its function. Street brings up an example of a child asking his parent whether he should confess to a prank that he committed and for which his friend has been wrongfully accused. Street argues that on Copp's (2008) view the parent might answer with something like "morally you should confess, and from a self-interested point of view you should stay silent", but that is not really answering the question. The child is asking what he should do *period*, and the parent would be forced to say that there is no answer to this question, only an answer to what to do from varying points of view. We have moral reasons, and we have self-grounded reasons, but one group does not outweigh the other. Given that moral deliberation is meant to be action guiding, resulting in some final reason for action all things considered, this deliberation, according to this response, seems to fail to *count* as moral deliberation. Furthermore, ordinarily, we think that if there is some moral requirement, then we ought to do that thing regardless of our self-interested reasons. So a moral naturalist theory that fails to be evaluatively realist, in the sense meant by Street, appears to fail to fulfil its function of guiding overall action.

However, the above would be a misreading of externalist views, and it is not even the way I think Street (2008) actually interprets those views. Externalist theories need not take the view that moral reasons and self-interested reasons are necessarily divergent. Copp (2009), for example, deals with the 'no reason *simpliciter'* objection by arguing that the default in evaluating deliberation is the standpoint of self-grounded reasons[39]. If Joyce (2006) is right that having a disposition to act morally generally leads to better outcomes long term than having the disposition to act from self-interest, then, as Frank (1987) shows, what might be the best thing to do in a given situation from a self-

[39] Recall from Chapter 1 that Copp (2009) attempts to deal with the reasons *simpliciter* objection by claiming that self-grounded reasons have 'default priority' in evaluating deliberation because such reasons are always relevant to evaluating deliberation, given what it is to deliberate. He therefore argues that "the default is to interpret the 'ought simpliciter' as the ought of practical rationality" (Copp, 2009; p. 36). Therefore, what reason we have to be moral, to endorse the moral system, will be derived from our self-grounded reasons.

interested perspective, is in fact to act from a moral perspective (or at least to give the moral perspective great weight). In the case of the prankster child, the question may not be "what should I do", but "what kind of person should I be". The answer to that question from both moral and self-interested perspectives, and thus the answer period, may well be to be the kind of person who acts from moral reasons, the kind of person who takes moral reasons to generally outweigh non-moral reasons.

An objection might be that the above approach may be self-contradictory and self-undermining. Joyce (2006) argues that to derive moral reasons from self-interested reasons rather than providing reasons independent of our desires is to risk undermining the bulwark against weakness of the will that morality provides. As soon as you start thinking in terms of self-interest, you are likely to be swayed by short-term self-interest. He also questions talk in terms of morality at all; why not talk simply in terms of desire and self-interest? And here we circle back around to the discussion of the role of moral discourse and whether an externalist-style revision can fulfil it. I take it that the discussion in Chapter 2 will suffice on this point, the short of it being that our psychological precommitments to morality are likely enough to ensure that the revised discourse can continue its function. Adopting FI-externalism could also help in this regard.

3.4.3. Street (2008) and Conceptual Indeterminacy

Aside from a possible argument about the function of morality, Street (2008) could also be read as making an argument that externalist theories fail to be realist about morality (rather than some 'schmorality') because they fail to be evaluatively realist. Even if most, perhaps all, people have reason *simpliciter* to act 'morally' on an externalist precisification, this fact is only contingent. Externalists must admit that an ideally coherent Caligula is not only possible, but also is not making any sort of mistake in holding their set of evaluative attitudes. Due to the desire-contingent nature of moral reasons on an externalist moral naturalist theory, we may be forced to conclude that an individual with a *wildly* divergent set of desires should act immorally. Consider an individual who not only has a strong desire to kill and little interest in acting morally, but also one who is indifferent to or seeks out their own death or imprisonment or other punishment, who cares little to not at all about living in a community with others, or achieving ends other than immoral ones. In this case, the externalist may have to conclude this individual should act 'immorally'.

A certain reading of Street (2008) might suggest that this is an anti-realist view, wherein because we are forced to say that this individual should act 'immorally', we may be forced to say this individual is doing nothing wrong in following through. Now this may not have much practical effect given ideally coherent Caligulas, if they exist, are few in number (it is highly doubtful that even the real Caligula was ever ideally coherent), and FI-externalism may allow us to live, justifiably, day-to-day 'make-believing' that such beings are doing something wrong, thus preserving our moral discourse in ordinary contexts. But this seems to still be a concern in our more critical contexts; it seems almost an abandonment of the concept of morality.

What this argument amounts to really, is an argument from our intuitions regarding what morality must be like: "it seems wrong for the Caligula to not be bound by morality, so any theory that suggests this is possible, must not actually be talking about morality". But as we know, this is the point of indeterminacy. Not everyone will share these intuitions, or think them reliable (especially considering they may be targeted by the EDA against evaluative realism), so this appears to be but one temperament among many that make up the point of indeterminacy. Assuming that morality is a natural kind[40], these intuitions would do little to resolve the indeterminacy. Our intuitions are the starting point for our questioning of the concept of morality: we have these intuitions about what morality is like, but nothing seems to satisfy all relevant properties, however there are imperfect claimants, is it acceptable for one of those to be morality? To use those intuitions to say 'no' seems to be begging the question at hand, but to use them to say 'yes' also seems too far. At most, these intuitions, if widely shared, may tell us what the original concept was, and thus what it would take to fully vindicate that concept. But they may also tell us what the imperfect claimants are, one of which may be sufficient to fulfil the same role.

As discussed in Chapter 1, the claim that externalist naturalism fails to count as morally realist because they fail to count as evaluatively realist (on Street's (2006, 2008) definition) is a problematic one. There are numerous ways of defining both moral realism and evaluative realism that will yield different answers as to whether a given externalist theory is realist or anti-realist. On Street's (2006) definition externalist theories are anti-realist, so might many theories in normative ethics we generally consider as being compatible with realism, such as preference utilitarianism (Berker, 2014). On the other hand, on Copp's (2009) definition of normative realism constructivist

[40] Admittedly, this is a fairly big assumption. However, considering that we are assuming we have available an externalist naturalist account that meets the Good-Reason constraint, then we are assuming that at least some parts of our concept of morality refer to some natural kind(s), such as facts about cooperation and human psychology.

theories are realist theories[41]. Joyce (2016a; p.27) meanwhile provides a definition of moral realism that also seems compatible with evaluative realism as Street understands it being false. It is hard to see how this brand of indeterminacy can be resolved in decisive favour of any particular definition of the realist/anti-realist divide in metaethics. The properties that each definition (or at least those above) tracks are all important to discussions in metaethics. A theory that satisfies Joyce's (2016a) definition but fails Street's (2008) is different from a theory that satisfies both or a theory that satisfies neither. For example, an externalist moral naturalist theory that claims that there are moral facts that hold independently of any of our evaluative attitudes but no desire-independent reasons for action is different from a theory that claims that the moral facts are relative to the individual. In some sense, the former is more realist than the latter.

3.4.4. The Moral Concept as a 'Mosaic'

One way of dealing with the indeterminacy in both the analytic definitions of morality and moral realism is through recognising that morality can be considered a 'mosaic' of different, and sometimes contradictory, elements, platitudes and functions, often with wildly varying genealogies (Sterelny and Fraser, 2016). Some of these elements will be debunked when faced with the EDA, such as beliefs about the practical clout of moral judgments, whereas others may be rescued, for example those moral beliefs that track truths about cooperation and the practices that support it. Joyce (2006) and Street (2008) make the case that without vindicating practical clout, the moral concept as a whole cannot be vindicated. But the mosaic nature of morality means that the options available to us are not merely full vindication or elimination of the concept of morality; instead partial vindication or revision is possible, vindicating some elements while eliminating or even revising others.

In fact, we have good reason to think that our folk concept of morality has continually undergone revisions throughout history. Sterelny and Fraser (2016) suggest that the "biological and cultural evolution of our moral practices very likely involved elements – norms of disgust, respect for authority, religion – that we now typically distinguish from moral thinking, properly so called…" (p. 4). The question is why is practical clout so special that to remove it is to abandon morality as a whole?

[41] At one point Copp (2009) even calls his pluralist-teleological view a "'constructivist' picture" (p.23).

As previously discussed, Joyce (2006) utilises the strategy of comparing the function of the proposed revised moral discourse and the original discourse in order to determine whether the revision of the concept is sufficient, but the mosaic nature of morality makes his utilisation of this strategy problematic. Joyce discusses only some functions of morality, that of promoting and signalling cooperative behaviour. There may well be other functions contained within the 'mosaic' of morality not included in his analysis. For example, in asserting the EDA and assuming that a compelling moral naturalist approach is not possible, Joyce ignores the tracking function of morality, thinking it unvindicated. But if we have reason to think that morality has the function of helping us track truths about our social environment, about cooperation and the practices that support it (Copp, 2008; 2009; Sterelny and Fraser, 2016), then we have reason to think that the truth-tracking function of morality is at least partially vindicated.

However, if Sterelny and Fraser (2016) are correct, there are likely other functions of morality as well. Ultimately it is an empirical matter what these may be, but the fact that there may be such other functions, limits our ability to say whether a revisionary approach (whether fictionalist, naturalist, constructivist etc.) is truly successful at fulfilling the function of the original discourse. The most we can say is that, in so far as morality evolved to track and promote cooperation and the social practices that support it, a revisionary approach (for example a FI-externalist approach) that meets the Good-Reason constraint likely could produce a discourse that can play that same role. And even this claim is dependent on empirical research. So on this methodology for resolving the point of indeterminacy surrounding whether an imperfect claimant is 'good enough', morality is only vindicated in so far as its function is to track and promote cooperation and the practices that support it. If morality really is a 'mosaic' of varying functions, then on this methodology, morality is only at most *partially* vindicated; some functions are vindicated, others are not or are yet to be. This may still be enough to help meet the epistemological challenge posed by the EDA, since the moral nativist hypothesis used in the EDA *is* that the evolved function of morality is the promotion of cooperation and the practices that support it; some moral beliefs appear to do this *because* they help us track truths about our environment.

On the flipside, even if it were true that a loss of perceived practical clout results in some impairment of the motivational function of moral discourse (which I argue in chapter 2 that this is not necessarily the case), then all we have at the moment is a *partial* debunking of morality. Other functions, such as the tracking of facts about cooperation, could well remain intact. If Copp (2009) is correct that morality, much like other normative systems such as etiquette, epistemic norms,

rationality etc., presents solutions to problems of 'normative governance', in this case solutions to coordination problems, then morality may well be able to play this role without practical clout. Ultimately it is an empirical matter whether this is the case, but it cannot just be dismissed out of hand. My point is that even if we can show that one function of morality is lost or impaired by the loss of practical clout in our moral concept, this does not show that all functions of morality are impaired or lost.

Even if we cannot fully vindicate the practical clout or the objective bindingness of morality on an externalist naturalist theory, it is not the case that we must eliminate practical clout; we could revise our conception of it. Street (2012) argues that "we have gone too far if we think that it is part of the very idea of morality that its requirements are categorical with respect to *any evaluative nature an agent might have*" (p. 18). Instead she argues "it is part of the very idea of morality that its requirements are categorical with respect to *some important parts of our evaluative nature*—for example, that it is categorical with respect to what we *desire* to do in an ordinary sense or what we find most appealing or pleasant" (p. 18). It may well be that our original concept of morality contained the former version of 'categorical', but the latter version does not seem too far off. Therefore, instead of arguing that to vindicate morality we must vindicate the objective bindingness of morality, we need only vindicate a kind of *relative* bindingness of morality, i.e. relative to our other evaluative attitudes. This bindingness is not separate from our evaluative nature, but part of it.

What Street (2012) suggests is needed to vindicate morality seems a lot like what I called our psychological precommitments to morality. Aspects of our psychology, our evaluative nature, that commit us to behaving in certain ways, according to certain requirements of a characteristic nature. This insight of Street's is important because it shows that what we want to vindicate, the perceived (objective) bindingness of morality, and what we have, bindingness relative to our other evaluative attitudes (for most humans), is not so great a gap. And if Joyce (2005) is right about the efficacy of fictionalism, the latter is just as functional as the former. An externalist theory, therefore, *can* vindicate the perceived bindingness of morality, it just happens to not be so 'objective'. This understanding helps make revising the moral discourse seem more permissible; even if we cannot vindicate the original moral concept completely because we cannot vindicate *inescapable authoritativeness*, an externalist theory can get most of the way there, vindicating *authority* of a kind i.e. over large aspects of our evaluative nature.

Acknowledging the mosaic nature of morality may help us with the indeterminacy surrounding the realism/antirealism divide. If morality is a mosaic of different elements, then we can be realist or antirealist about different elements separately. If, for example, an otherwise compelling externalist theory can be provided that appears to meet the EDA, we may be able to be realists about the moral facts and the natural-moral identity, believing that they hold independently of any of our evaluative attitudes, while being anti-realist about their normativity, believing that they provide no reasons (*simpliciter*) independent of any of our evaluative attitudes. I have the temperament that we might then say that such a theory is morally realist yet evaluatively anti-realist, but labelling in this way is not really necessary so long as it is clear what exactly we are being realists or anti-realists about, namely the moral facts and their natural-moral identity, and their normativity, respectively. As the difference between moral fictionalism and FI-externalism shows, we may take fictionalist attitudes to some elements but not others. The moral fictionalist takes a fictionalist attitude toward many elements and an abolitionist approach to others, while the FI-externalist takes a fictionalist attitude only to practical clout and the other elements are kept or eliminated according to the revision. The point is that, again, we need not take an all-or-nothing approach to the mosaic of morality; we can separate its component elements out and classify a theory's position on each separately[42].

3.4.5. The Partial Vindication of Moral Discourse

If we have an externalist naturalist theory that meets the Good-Reason constraint, it appears we have an answer for whether morality is vindicated. That answer is that we appear to have a *partial* vindication of morality, where some of the functions and platitudes of the 'mosaic' of the moral concept are (at least partially) vindicated and others are not. But we still have not reached an answer to the question of '*can* we continue to use the term 'moral'? Can the imperfect claimant suggested by the externalist naturalist theory claim the prize? While morality may be partially vindicated, it is equally partially debunked. Perhaps the fact that it is partially vindicated gives us some allowance to keep using the term 'moral', but perhaps the fact that it is partially debunked gives the error theorist allowance to eliminate it. In some ways this is similar to a view advocated by Joyce (2016c); metaethical ambivalence:

[42] As an aside, it seems like Street takes an approach more similar to this in her later works. Street (2012) could be read as hinting toward an externalist moral naturalist theory in discussing the 'characteristic content' of moral requirements, the content of certain judgments that makes them judgments about the *moral* thing to do.

> This perspective begins with a kind of metametaethical enlightenment. The moral naturalist espouses moral naturalism, but this espousal reflects a mature decision, by which I mean that the moral naturalist doesn't claim to have latched on to an incontrovertible realm of moral facts of which the skeptic is foolishly ignorant, but rather acknowledges that this moral naturalism has been achieved only via a nonmandatory piece of conceptual precisification. (This describes Lewis's tolerant view.) Likewise, the moral skeptic champions moral skepticism, but this too is a sophisticated verdict: not the simple declaration that there are no moral values and that the naturalist is gullibly uncritical, but rather a decision that recognizes that this skepticism has been earned only by making certain non-obligatory but permissible conceptual clarifications. (Joyce, 2016c, p. 105)

In addition, Joyce (2016c) advocates not mere grudging acceptance that the opposition is warranted in their views, but a willingness to sometimes adopt the other position in order to gain the insights and benefits of that view. This then is also a view reached on pragmatic grounds, not just epistemic. Joyce (2016c) attempts to weigh up the pragmatic benefits of either view to determine whether adopting one view is better than the other. He recognises that the moral naturalist may have some benefits, but argues that it is a mistake to think that the moral error theory does not or that everyone would prefer the benefits of the naturalist view to the sceptical.

If we therefore have allowance for either revision or scepticism, then the question is 'should we revise or should we eliminate'? This is a pragmatic question, meant to be answered on the basis of benefits of either approach. The sceptic might ask that if we only have a partial vindication, why do we need to use the term 'moral', why not talk in terms of desires and beliefs? The answer, although ultimately this is an empirical matter, likely lies in our psychological precommitments to moral discourse– as argued in Chapter 2 our moral precommitments appear to provide us with numerous benefits – dropping 'morality' likely means losing those benefits. Joyce (2005, 2006) argued that the sceptical approach can still achieve at least some of the benefits of morality by adopting moral fictionalism. I argued that the moral naturalist should be able to achieve those same benefits by adopting FI-externalism. On the reasonable assumption that it is better to proceed from attitudes of belief in what is true than proceed from attitudes of make-belief in what is false, that is it is best to limit fictionalist attitudes as much as possible, then FI-externalism seems preferable.

However, Joyce (2016c) introduces another benefit of a sceptical approach; that there is a benefit to being epistemologically shaken, to finding out that we are wrong about something that seems so fundamental, and instead finding out 'how mysterious everything really is' (p. 102). He argues that "[i]t is both a corrective to epistemic complacency and a spur to intense reflection and inquiry" (Joyce, 2016c; p. 102). Now this seems a rather minor and nebulous benefit to me, one that does not seem like it would outweigh the costs of holding a fictionalist attitude or losing moral discourse, but Joyce also makes an important point about the desire-contingent nature of our pragmatic reasons. It is certainly conceivable that an individual may vastly prefer this benefit to any benefit that comes from moral discourse, so we cannot say for sure which view is better for any given person.

But the question of whether to revise our moral concept or to abandon it is not about individuals, it is a collective decision. And in making his case for a revisionary moral fictionalism, Joyce (2005) seems to agree:

> Let us just say when morality is removed from the picture, what is practically called for is a matter of a cost-benefit analysis, where the costs and benefits can be understood liberally as preference satisfaction. By asking what *we* ought to do I am asking how a *group* of persons, who share a variety of broad interests, projects, ends – and who have come to the realization that morality is a bankrupt theory – might best carry on. (Joyce, 2005; p. 288)

Now in our case we are not assuming that morality *is* a bankrupt theory, our assumption is that we have an externalist theory that meets the Good-Reason constraint and shows that morality is partially vindicated. The point is that the question of what to do with our concept of morality is a collective one, and even if proper convergence is not assured, there is nevertheless a rough kind of convergence wherein we, as human beings, share a variety of broad interests, projects and ends. The matter is ultimately an empirical one, but it seems we should be able to meet many of these interests, projects and ends through utilisation of a revised moral discourse.

If this is the case, that we, as a society, have strong pragmatic reasons to revise our moral discourse, and we have an externalist theory that meets the Good-Reason constraint, then it certainly seems that a revision of our moral discourse is permissible. Joyce (2006) and Street's (2008) arguments against the likelihood of a compelling moral naturalist theory therefore fall through. Provided an externalist theory can meet the 'trivially question-begging' objection (which seems quite

possible, if not likely), and the empirical matter of the functionality of the discourse is solved in the revisionists favour, then an externalist naturalist theory should be able to count as a theory about morality, meeting the Desiderata constraint. Even if the empirical matter is unsolved or remains so, there is no reason to think that an externalist theory that can meet the Good-Reason constraint is *unlikely* to be compelling. What's more is that such a theory could be considered a *realist* theory of morality even if it were not evaluatively realist. Street (2008) is quite likely right that externalist theories are compatible with constructivist theories of normativity, but this does not make them any less *morally* realist.

Chapter 3 Conclusion

In this chapter, I discussed whether moral naturalist theories can resist Street's (2006) EDA, the Darwinian Dilemma. First, I discussed Street's original Darwinian Dilemma, the naturalist response, and Street's two objections to the naturalist. I then explored whether an externalist theory could resist Street's Darwinian Dilemma, and her 'trivially question-begging' and 'one level up' objections. I showed that it can, but only by meeting the Good-Reason constraint and only at the expense of failing to be evaluatively realist. I then discussed the arguments made by Joyce (2006) and Street (2006) that externalist naturalist theories fail to be theories about morality, concluding that an externalist naturalist theory that meets the Good-Reason constraint should be able to count as a theory about morality because it can at least partially vindicate morality.

Conclusion

Evolutionary debunking arguments (EDAs) that target morality aim to undermine the justification of our moral beliefs by arguing that such moral beliefs are likely the result of evolutionary forces that are insensitive to the truth. For Joyce (2006), the moral nativist hypothesis suggests that the emergence and persistence of the relevant faculty can be wholly explained by non-truth-tracking functions, so unless it can be shown that the moral truth plays an explanatory role in our moral belief-formation process, we have no justification to think that that our moral beliefs are true. For Street (2006) the argument is presented as the 'Darwinian Dilemma', where evaluative realists must either assert or deny that there is some relation between our evolutionarily influenced evaluative attitudes and the independent evaluative facts. If we deny that such a relation exists, then, given the tremendous number of possible evaluative beliefs we could have, we would be committed to an 'implausibly large coincidence' that we just happened to land on the right ones. If we assert that there is such a relation, then we are challenged with providing an explanation of what that relation is. Street argues that the realist must accept some version of the 'tracking account', that we evolved to have the evaluative beliefs we do because they were true and it was fitness enhancing to perceive this fact. However, she argues that the account is scientifically inferior to a different explanation, the 'adaptive link account', where we came to hold such evaluative beliefs because it was fitness enhancing to hold them regardless of whether they were true or not. She therefore argues that we should accept the adaptive link account over the tracking account, undermining the justification in our evaluative beliefs. The arguments of Joyce (2006) and Street (2006) are therefore similar in that they both pose a challenge to the moral success theorist to provide a *compelling* explanation of how our moral beliefs evolved to track the truth. One way this has been attempted is through moral naturalism, identifying or grounding the moral facts in some set of natural facts that appear to play an explanatory role in our belief-formation process.

Both Joyce (2006) and Street (2006; 2008) argue that no extant moral naturalist theory is compelling enough to meet the challenge posed by the EDA, and furthermore both make arguments that suggest that providing such a compelling moral naturalist account is highly unlikely, if not impossible. Joyce argues that it is a requirement for a theory to be about morality that it accounts for the perceived *inescapable authoritativeness*, or practical clout, of morality. Yet, he argues, no moral naturalist theory can satisfactorily account for practical clout. Street argues that value naturalism is subject to the same Darwinian Dilemma 'one level up' and is 'trivially question-

begging', relying on the very evaluative attitudes whose reliability is at stake to establish the natural-evaluative identity. Street (2008) can also be read as making a similar argument as Joyce (2006), arguing that for a naturalist theory to be a realist theory of *morality*, it must be evaluatively realist as well, holding that our reasons for action hold independently of any of our evaluative attitudes. Revisionary externalist approaches however, argue that it is *not* necessary for naturalist theories to account for practical clout to be considered theories about *morality*. They argue that although there may not be any property in the world that can fully satisfy our ordinary concept of morality, we do not need it to. If an imperfect claimant can satisfy most of what we want then that claimant may be 'good enough' to assume the term 'morality'.

Determining whether any revisionary approach is 'good enough' to count as *morality* is a difficult prospect. Joyce (2006, 2016c) suggests we may have a chance of resolving this indeterminacy by examining what the concept is *used for*, its function, and determining whether the revised discourse can play the same role. Joyce (2006) argues that a morality without practical clout would be unable to fulfil its function; that of providing three major benefits that support and enable cooperation, the benefits to personal commitments, dyadic commitments and social commitments. However, elsewhere in his work, in his argument for moral fictionalism, we can find the resources needed to defend against his argument against externalist moral naturalism, establishing that externalist moral naturalism should actually be able to achieve at least the first two benefits.

In order to show how a revisionary discourse can capture the last benefit, I introduce fictional-internalist externalism, or FI-externalism for short. FI-externalism holds that that there are moral facts that do not provide desire-independent reasons for action, but, in ordinary contexts, we can *make-believe* they provide desire-independent reasons for action. Adopting FI-externalism would allow one to achieve all three suggested benefits of morality, provided moral fictionalism can as well. Furthermore, moral fictionalism tends to be unstable, fictions tend to go astray without strict discipline and the criteria needed to keep them focused on meeting its function. So FI-externalism has the benefit of limiting the fiction, and thus limiting the source of instability. According to Joyce's methodology then, an otherwise compelling FI-externalist moral discourse should be capable of fulfilling the same role as the old moral discourse, and thus the revisionary externalist theory in question should be 'good enough' to count as a theory about *morality.*

The externalist can also plausibly deal with Street's (2006, 2008b) objections. The 'trivially question-begging' objection can be dealt with by recognising that it really is just another version of

the challenge of providing a *compelling* moral naturalist account and not merely a *possible* one. That is to say, it pushes the naturalist to provide an actual account of how and why the moral facts are identical to, or grounded by, the natural facts posed by their theory, and thus how our moral beliefs, at least partially, tracked the independent moral truth. To satisfy the challenge the naturalist must provide good reason, independent of their moral beliefs and intuitions, to accept the natural-moral identity. I outline an externalist moral naturalist theory introduced by Sterelny and Fraser (2016) that appears to meet these criteria, providing actual empirical evidence for their theory instead of presenting it as a mere possibility. I argue that is entirely plausible that an externalist theory along lines of Sterelny and Fraser (2016) could meet the challenge of the 'trivally question begging' objection. In regards to Street's (2006, 2008) 'one level up' objection, I argue that the externalist moral naturalist can avoid the force of the original Darwinian Dilemma that targets the evaluative domain by taking the 'rigidifying move', fixing the truth of the natural-moral identity according to our *actual* evaluative judgments. Taking this approach means an otherwise compelling moral naturalist theory can vindicate moral realism, but only at the expense of failing to vindicate evaluative realism.

Another argument that can be read from Street (2008), that for a theory to be a realist theory about *morality*, it must be evaluatively realist, is very similar to Joyce's (2006) requirement of practical clout. The argument can be dealt with in similar ways to how we dealt with Joyce's argument. If the argument is about the function of moral discourse, then we should be able show that an externalist moral naturalist theory should be satisfactory if it is otherwise compelling. If the argument is more than just about the function of morality, than we can deal with the indeterminacy in other ways. We can, for instance, recognise that the ordinary folk concept of morality is a 'mosaic' of different functions and platitudes (Sterelny and Fraser, 2016), of which truth-tracking is only one function. Even if we have good reason to accept an externalist moral naturalist theory in response to the EDA, this is only a partial vindication of morality; many moral beliefs and platitudes will not serve this truth-tracking function. We would thus likely need to abandon many of our folk moral beliefs and platitudes, including our intuitions that moral facts provide reasons to act accordingly regardless of our evaluative attitudes. But the choice is not merely one of full vindication or elimination; we can vindicate some beliefs and eliminate others. This partial vindication gives us license to adopt a revisionary approach, and since we have good reason to think the practical benefits of a revisionary approach are greater than the benefits of an error theoretic approach (even when adopting moral fictionalism), then we have good reason to adopt the revision. Recognising the mosaic nature of morality, also helps with dealing with the question of whether an externalist theory is a realist or

anti-realist theory or morality; such a theory may fail to be evaluatively realist, or fail to be realist about practical clout, but it *is* realist about other the truth-tracking aspects of morality.

It may be the case then that, on the question of whether moral realism is vindicated or eliminated in response to the EDA, the answer may be somewhere in the middle. It may well be that internalist definitions most appropriately fit our ordinary usage, but in their failure to defend against the Darwinian Dilemma and other EDAs, externalist reductions may be 'good enough', leading to a kind of 'demi-realism' about morality where there are no stance-independent moral facts that provide stance-independent reasons for action, but there are stance-independent moral facts that provide reasons for action that are dependent on our evaluative attitudes. Alternatively, we might take an approach similar to Street (2012), and revise our concept of practical clout to be authoritative with respect *to some important parts of our evaluative nature*, for example with respect to what we find most appealing or pleasant. Moral realism, on either view, would therefore be only partially vindicated in response to the EDA.

Joyce (2006) and Street (2006, 2008) have both made arguments that moral naturalist theories are not compelling enough, and likely will never be compelling enough, to meet the epistemological challenge of their respective EDAs. In this thesis, I have made the argument that externalist moral naturalist theories, provided they are otherwise compelling, can largely avoid the force of their arguments regarding the equivocally analytic definition of morality. In addition, I have argued that we have some reason to think that an otherwise compelling externalist moral naturalist theory could be provided, if it has not already. That is to say, that as long as an externalist account that is not merely possible can be provided, which it plausibly can, then we can successfully meet the epistemological challenge of the EDA by taking an externalist revisionary approach to morality.

References

Alexander, R. (1987). *The Biology of Moral Systems*. New York: Routledge

Bagnoli, C. (2017). Constructivism in Metaethics. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Retrieved from: https://plato.stanford.edu/archives/win2017/entries/constructivism-metaethics

Bedke, M. (2009). Intuitive Non-Naturalism Meets Cosmic Coincidence. *Pacific Philosophical Quarterly, 90*, pp. 188–209.

Berker, S. (2014). Does Evolutionary Psychology Show That Normativity Is Mind-Dependent? In J. D'Arms and D. Jacobson (eds.), *Moral Psychology and Human Agency: Philosophical Essays on the Science of Ethics* (pp. 215–52). Oxford: Oxford University Press

Brink, D.O. (1986). Externalist Moral Realism. *Southern Journal of Philosophy, 24*(1), pp. 23-41. Doi: 10.1111/j.2041-6962.1986.tb01594.x

Boyd, R. (1988). How to be a Moral Realist. In G. Sayre-McCord (ed.) *Essays on Moral Realism* (pp. 181-228). New York: Cornell University Press.

Clarke-Doane, J. (2012). Morality and Mathematics: The Evolutionary Challenge. *Ethics*, *122*(2), pp. 313–40.

Copp, D. (2008). Darwinian Skepticism about Moral Realism. *Philosophical Issues, 18,* pp. 186-206. doi: 10.1111/j.1533-6077.2008.00144.x

Copp, D. (2009). Toward a Pluralist and Teleological Theory of Normativity. *Philosophical Issues, 19*(1), pp 21-37. doi: 10.1111/j.1533-6077.2009.00157.x

Copp, D. (2012). Varieties of Moral Naturalism. *Filosofia UNISINOS, 13*(2), pp. 280-295. doi: 10.4013/fsu.2012.132(suppl).05

Das, R. (2016). Evolutionary debunking of morality: epistemological or metaphysical? *Philosophical Studies, 173,* pp. 417-435. doi: 10.1007/s11098-015-0499-9

Darwall, S. (1997). Reasons, Motives, and the Demands of Morality: An Introduction. In S. Darwall, A. Gibbard, & P. Railton (eds.), *Moral Discourse and Practice:Some Philosophical Approaches.* New York: Oxford University Press

Dennet, D. C. (1995). *Darwin's Dangerous Idea.* New York: Simon and Schuster

FitzPatrick, W. J. (2014).'Why There is No Darwinian Dilemma for Ethical Realism'. In M. Bergmann & P. Kain (eds.), *Challenges to Moral and Religious Belief: Disagreement and Evolution*. Oxford Scholarship Online. doi: 10.1093/acprof:oso/9780199669776.001.0001

Frank, R. H. (1987). If Homo Economicus Could Choose His Own Utility Function, Would He Want One with a Conscience? *The American Economic Review, 77*(4), pp. 593-604.

Frank, R. H. (1988). *Passions within Reason: The Strategic Role of the Emotions.* New York: Norton

Gendler, T. S., and Hawthorne, J. (2002). Introduction: Conceivability and Possibility. In T. S. Gendler, & J. Hawthorne (eds.) *Conceivability and Possibility*. Oxford: Clarendon Press

Greene, J. (2008). 'The Secret Joke of Kant's Soul.' In W. Sinnot-Armstrong (ed.), *Moral Psychology, Vol. 3: The Neuroscience of Morality, Emotions, Brain Disorders and Development.* (pp.35-80) Cambridge, MA: MIT Press

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

Harman, G. (1986). Moral Explanations of Natural Facts: Can moral claims be tested against moral reality? *The Southern Journal of Philosophy, 24*(5), pp. 57-68. Retrieved from: https://search-proquest-com.ezproxy.lib.monash.edu.au/docview/1307504951?accountid=12528

Husi, S. (2014). Against Moral Fictionalism. *Journal of Moral Philosophy, 11,* pp. 80-96. doi: 10.1163/17455243-4681008

Jordan, J. J., Sommers, R., Bloom, P., and Rand, D. G. (2017). Why Do We Hate Hypocrites? Evidence for a Theory of False Signaling. *Psychological Science, 28*(3), pp. 356-368. doi: 10.1177/0956797616685771

Joyce, R. (2005). 'Moral Fictionalism', in M.E. Kalderon (ed.), *Fictionalism in Metaphysics*. Oxford: Oxford University Press, pp. 287-313

Joyce, R. (2006). *The Evolution of Morality.* Cambridge: Mit Press

Joyce, R. (2009). Is Moral Projectivism Empirically Tractable? *Ethical Theory and Moral Practice, 12*, pp. 53-75. doi: 10.1007/s10677-008-9127-5

Joyce, R. (2016a). Reply: Confessions of a Modest Debunker. In U. D. Leibowitz & N. Sinclair (eds.), *Explanation in Ethics and Mathematics: Debunking and Dispensibility* (pp. 124-144). doi: 10.1093/acprof:oso/9780198778592.003.0007

Joyce, R. (2016b). Morality, Schmorality. In *Essays in Moral Skepticism* (pp. 41-66). Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780198754879.003.0003

Joyce, R. (2016c). Metaethical Pluralism. In *Essays in Moral Skepticism* (pp. 89-106). Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780198754879.003.0005

Joyce, R. (2016d). Evolution, Truth-tracking, and Moral Skepticism. In *Essays in Moral Skepticism* (pp. 142-157). Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780198754879.001.0001

Joyce, R. (2016e). Irealism and the Genealogy of Morals. In *Essays in Moral Skepticism* (pp. 159-174). Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780198754879.003.0009

Kahane, G. (2011). Evolutionary Debunking Arguments. *Noûs, 45*(1), pp. 103-125. doi: 10.1111/j.1468-0068.2010.00770.x

Kalderon, M. E. (2005). *Moral Fictionalism.* Oxford: Clarendon Press

Kitcher, P. (2011). *The ethical project*. Cambridge, MA: Harvard University Press

Killerby, C. K. (2002). *Sumptuary Law in Italy 1200-1500*. New York: Oxford University Press. Retrieved from: https://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199247936.001.0001/acprof-9780199247936.

Korsgaard, C. (1996). *The Sources of Normativity.* New York: Cambridge University Press

Mackie, J. L. (1977). *Ethics: Inventing right and wrong.* Harmondsworth: Penguin

Ruse, M. (1986). *Taking Darwin seriously*. Oxford: Basil Blackwell.

Ruse, M. (2006). *Darwinism and its discontents*. Cambridge: Cambridge University Press.

Shafer-Landau, R. (2012) Evolutionary Debunking, Moral Realism and Moral Knowledge. *Journal of Ethics & Social Philosopy, 7*(1), pp. 1-37. Retrieved from: http://link.galegroup.com/apps/doc/A323259253/AONE?u=monash&sid=AONE&xid=8946e564.

Singer, Peter. (2005) Ethics and Intuitions. *The Journal of Ethics*, *9*(3), pp. 331–352. doi: 10.1007/s10892-005-3508-y

Skitka, L. J., Bauman, C. W., and Sargis, E. G. (2005) Moral conviction: Another contributor to attitude strength or something more? Journal of Personality and Social Psychology, 8, pp. 895-917. doi: 10.1037/0022-3514.88.6.895

Smith, M. (1994). *The Moral Problem.* Oxford: Blackwell

Smith, M., Lewis, D., and Johnston, M. (1989). Dispositional Theories of Value. *Proceedings of the Aristotelian Society, Supplementary Volumes, 63,* pp. 89-174. Retrieved from: https://www-jstor-org.ezproxy.lib.monash.edu.au/stable/4106918

Sober, E., and Wilson, D. S. (1998) *Unto Others: The Evolution and Psychology of Unselfish Behavior.* Cambridge, Massachusetts: Harvard University Press

Stanford, P. K. (2018). The Difference Between Ice Cream and Nazis: Moral Externalization and the Evolution of Human Cooperation. *Behavioral and Brain Sciences 41,* pp. 1-57. doi: 10.1017/S0140525X17001911

Sterelny, K., & Fraser, B. (2016). Evolution and Moral Realism. *The British Journal for the Philosophy of Science, 0,* pp. 1-26. doi: 10.1093/bjps/axv060

Stevenson, C. L. (1937). The emotive meaning of ethical terms. *Mind*, *46,* pp. 14–31. doi: 10.1093/mind/XLVI.181.14

Street, S. (2006). A Darwinian Dilemma for Realist Theories of Value. *Philosophical Studies, 127*(1), pp. 109-166. doi: 10.1007/s11098- 005-1726-6

Street, S. (2008). Reply to Copp: Naturalism, Normativity, and the Varieties of Realism Worth Worrying About. *Philosophical Issues, 18,* pp. 207-228. doi: 10.1111/j.1533-6077.2008.00145.x

Street, S. (2009). Evolution and the Normativity of Epistemic Reasons. *Canadian Journal of Philosophy, 39*(1), p. 213-248. doi: 10.1080/00455091.2009.10717649

Street, S. (2012). Coming to Terms with Contingency: Humean Constructivism about Practical Reason. In J. Lenman & Y. Shemmer (eds.), *Constructivism in Practical Philosophy* (pp. 40-59). Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199609833.001.0001

Sturgeon, N. (1985). Moral Explanations. In D. Copp & D. Zimmerman (eds.) *Morality, Reason, and Truth.* (p. 49-78). Totowa, Rowman and Allanheld

Tropman, E. (2014). Evolutionary debunking arguments: moral realism, constructivism, and explaining moral knowledge. *Philosophical Explorations, 17*(2), pp. 126-140. doi: 10.1080/13869795.2013.855807

Vavova, K. (2014). 'Debunking Evolutionary Debunking.' In R. Shafer-Landau (ed.), *Oxford Studies in Metaethics, Vol. 9,* (pp. 76-101). Oxford: Oxford University Press

Vavova, K. (2015). Evolutionary Debunking of Moral Realism. *Philosophy Compass, 10*(2), pp. 104-116. doi: 10.1111/phc3.12194

Williams, B. (1981). 'Internal and External Reasons'. In B. Williams, *Moral Luck Philosophical Papers, Oxford Studies in Metaethics volume 13 1973-1980.* Cambridge, Cambridge University Press, (pp. 101-113). Doi: 10.1017/CBO9781139165860.009

Appendix A: Glossary of Key Terms

Evaluative Attitude: From Street (2006): "*Evaluative attitudes* I understand to include states such as desires, attitudes of approval and disapproval, unreflective evaluative tendencies such as the tendency to experience X as counting in favor of or demanding Y, and consciously or unconsciously held evaluative judgements, such as judgements about what is a reason for what, about what one should or ought to do, about what is good, valuable, or worthwhile, about what is morally right or wrong, and so on" (pg. 110).

Evaluative/normative Fact: From Street (2006): "*Evaluative facts* or *truths* I understand as facts or truths of the form that X is a normative reason to Y, that one should or ought to X, that X is good, valuable, or worthwhile, that X is morally right or wrong, and so on" (pg. 110).

N.B: 'Evaluative' and 'normative' are often used interchangeably in the literature; 'evaluative' tends to put the focus on 'these are facts about values' and 'normative' tends to put the focus on 'these are facts about what reasons we have for action'.
Also note: Moral facts are generally considered to be a subset of evaluative/normative facts

Evaluative/normative realism: There are at least some evaluative facts that hold independently of any and all of our evaluative attitudes.

Two varieties:
Naturalist evaluative realism (aka value naturalism): There are at least some evaluative facts that hold independently of any and all our evaluative attitudes, and all of these evaluative facts are either identical to or entirely grounded in natural facts. (Berker, 2014)
Non-naturalist evaluative realism: There are at least some evaluative facts that hold independently of all our evaluative attitudes, and at least some of these evaluative facts are non-natural and not grounded in natural facts

Its negation:

Evaluative/normative antirealism: There are no evaluative facts that hold independently of any and all of our evaluative attitudes.

Two varieties:

Nihilist evaluative antirealism: There are no evaluative facts.

Non-nihilist evaluative antirealism: There are at least some evaluative facts, and all of these evaluative facts are at least partially grounded in facts about our evaluative attitudes (e.g. **Humean Constructivism**)

Evolutionary Debunking Argument (EDA): arguments which attempt to undermine our justification in believing a particular belief or set of beliefs from the fact that the belief or set of beliefs in question are the result of evolutionary processes.

Darwinian Dilemma: The Evolutionary Debunking Argument introduced by Street (2006) that targets evaluative realism. Poses a dilemma to the evaluative realist, with both horns leading to a sceptical conclusion.

Moral Cognitivism: The view that moral sentences express propositions.

Moral Error Theory: The view that our moral sentences are systematically in error (i.e. false), either because there are no moral facts, or because no moral judgments are epistemically justified.

Its negation:

Moral Non-cognitivism: The view that moral sentences do not express propositions and therefore are not truth-apt

Moral (Reasons) Externalism: The view that moral facts may provide reasons for action, but not independently of any and all of our evaluative attitudes.

Moral (Reasons) Internalism: The view that moral facts provide reasons for action independent of any and all of our evaluative attitudes. May include the view that moral judgments hold **practical clout** (see below). N.B. Moral reasons internalism should be distinguished from

moral motivational internalism that argues that moral judgments always come with a motivation to act accordingly.

Throughout this thesis when I refer to 'internalism' I will be referring to reasons internalism not motivational internalism.

Moral Realism: There are at least some moral facts that hold independently of any and all of our evaluative attitudes.

Two varieties:

Moral Naturalism: There are at least some moral facts that hold independently of any and all our evaluative attitudes, and all of these moral facts are either identical to or entirely grounded in natural facts.

Non-naturalist moral realism: There are at least some moral facts that hold independently of all our evaluative attitudes, and at least some of these moral facts are non-natural and not grounded in natural facts

Its negation:

Moral Antirealism: There are no evaluative facts that hold independently of any and all of our evaluative attitudes.

Varieties:

Moral Nihilism: There are no moral facts.

Moral Abolitionism: We should not use moral terms as they are false or do not refer to anything

Moral Fictionalism: We should treat morality as a 'useful fiction', accepting that there are no moral facts in our most critical contexts (such as the philosophy classroom, etc.) while 'make-believing' in moral facts and continuing to use moral discourse in our ordinary contexts (everyday life), generally in order to continue to receive the benefits of moral discourse

Non-nihilist moral antirealism: There are at least some moral facts, and all of these moral facts are at least partially grounded in facts about our evaluative attitudes (e.g. Relativism, Subjectivism, non-cognitivism, **Humean Constructivism** etc.)

Practical Clout: Moral judgments that are seen to hold practical clout are seen to be *inescapably authoritative.* If moral judgments are seen to be inescapable, they are seen to apply to everyone everywhere. If moral judgments are seen to be authoritative, they are seen to provide reason for action independent of any and all of our desires.

Precommitment: The set of thought patterns and psychological mechanisms, instilled in us whether through genetics, our upbringing or both, that push us towards certain kinds of action, thus serving in a similar capacity to a commitment device (Frank, 1988). In effect, to hold a 'precommitment' to some set of actions, is to be 'previously committed' to performing those kinds of actions, it is not that one actively decides to commit to those actions in the moment, or that one decides to commit to perform such actions in the future, but, by virtue of one's psychology, one has certain future options cut off (or at least made much less appealing).

Appendix B: Joyce's (2006) arguments in Standard Form

Argument for (agnostic) error theory (Joyce, 2006)

P1. Unless we have a compelling account for moral naturalism, the Evolutionary Debunking Argument for morality gives us good reason to suspect that our moral beliefs are unjustified

P2. In order for a moral naturalist account to be compelling, it must either take a vindicatory approach (first horn) or a revisionary approach (second horn)

First horn:

P3. In order for a moral naturalist account to take a vindicatory approach, it must explain how moral judgments can have practical clout

P4. For a moral naturalist account to explain how moral judgments can have practical clout, it must be possible to show how naturalistic facts can provide practical clout

P5. It is impossible to show how naturalistic facts can provide practical clout

C1. It is impossible for a moral naturalist account to take a vindicatory approach

Second horn:

P6. In order for a moral naturalist account to take a revisionary approach, it must show how its proposed revised discourse can play the same role (fulfil the same function), as the original discourse, to act as a bulwark against weakness of the will and to coordinate social interaction

P7. In order for a proposed revised discourse to act as a bulwark against weakness of the will and to coordinate social interaction, moral judgments in the proposed revised discourse must be seen as holding practical clout

P8. In order for moral judgments in the proposed revised discourse to be seen as holding practical clout, the moral judgments in the proposed revised discourse must hold practical clout

P9. Moral judgments in the proposed revised moral discourse do not hold practical clout

C2. It is impossible for a moral naturalist account to take a revisionary approach

C3. It is impossible to provide a compelling moral naturalist account

C4. We have good reason to suspect that our moral beliefs are unjustified

Internalist moral naturalist theories would attempt to deny P5. Externalist moral naturalist theories would likely attempt to deny P7.

Argument for moral fictionalism (Joyce, 2005, 2006)

P1. We have good reason to suspect that our moral beliefs are unjustified and we generally should not believe things we have no justification for

P2. However, moral discourse is useful as it provides regulative benefits, acting as a bulwark against weakness of the will and coordinating social interaction

P3. We should (in the prudential sense of "should", not the moral sense) try to keep these regulative benefits

P4. We can obtain the regulative benefits from moral discourse without believing in it as being literally true, by adopting moral fictionalism and making use of a precommitment to morality

P5. We already generally have a psychological precommitment to moral discourse

C1. We ought to adopt moral fictionalism