HOW NOT TO RESPOND TO EVOLUTIONARY DEBUNKING ARGUMENTS

Christoph Lernpaß

THIS THESIS IS SUBMITTED IN FULFILMENT OF THE DEGREE OF MASTER OF ARTS (PHILOSOPHY) AT THE UNIVERSITY OF OTAGO, DUNEDIN, NEW ZEALAND

AUGUST 2018

ABSTRACT

Evolutionary debunking arguments [EDAs] in moral epistemology generally aim to show that what we know or what we have good reason to believe about the impact of evolution on the development of human moral psychology threatens the epistemic standing of our moral beliefs, and therefore we have a *prima facie* reason to withhold judgment in moral matters. Very often though, this claim is qualified to target only meta-ethical theories of a certain stripe. The prime target here has been a strong kind of moral realism. In those cases, it is argued that the sceptical conclusion would only follow if we assume strong moral realism. In turn, this is then taken to be a strong reason for rejecting those theories targeted by the respective EDAs.

In this thesis, I will look specifically at certain replies to EDAs that have been developed on behalf of a strong moral realism. I will argue that the two most prominent members of a family of popular responses to EDAs, which I call the "standard responses" (in virtue of their popularity), are ill-suited for neutralizing the epistemic threat that supposedly arises from evolutionary considerations.

I will argue that these two standard responses cannot make good on their promise that the epistemic threat supposedly arising from evolution can be neutralized, even if the debunkers' empirical story is largely correct, and if we assume a strong moral realism.

It is a plausible desideratum on any satisfying response to EDAs that the response should support the claim that evolutionary considerations do not show that our moral beliefs are seriously epistemically deficient or that we are seriously epistemically deficient for holding them. In other words: a good response to the EDAs should show that evolutionary considerations do not suffice to render us epistemically criticisable for holding or continuing to hold our moral beliefs. I will argue that the considerations offered by the two standard responses are insufficient for satisfying this desideratum.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my primary supervisor, Alex Miller, for his tireless dedication in seeing this project through, for both his attention to detail and his keen eye for the larger picture and for helpful feedback and discussion. It is no overstatement to say that this project would have been a perfect mess without his support. Any remaining messiness is of course entirely my own fault.

My second supervisor, Greg Dawes, also deserves special mention for going well beyond what one could reasonably expect in terms of providing invaluable feedback and discussion.

Thanks to Robin McKenna, whose excellent course "Moral Anti-Realism" (Winter 2016) at the University of Vienna first raised my interest in evolutionary debunking arguments in moral epistemology. Thanks to James Maclaurin for discussing basic philosophy of biology-questions with me in the early stages of my research.

Thanks to the audience of my presentation of material from this thesis at the 2018 NZAP/AAP-conference in Wellington for their questions and feedback. Many thanks to staff and postgrads of the Department of Philosophy at Otago for providing quite beneficial feedback to my presentations of material from this thesis (at various stages of development) at the departmental postgraduate-conference and at three sessions of the weekly postgraduate seminar. Special thanks to Chris Lean and Charles Pidgen for good advice, keen observations and for asking the hard-hitting questions.

Thanks to the Department of Philosophy at Otago as a whole for being friendly and supportive throughout my stay in New Zealand. Special thanks to Kirk Michaelian and to Chloe Wall for all-around marvelousness (that's a real word!) in helping me to get set up and orient myself in Dunedin and around the University. Thanks to Tiddy Smith for providing feedback on parts of my thesis, and for answering countless questions about how to best phrase/format things. Thanks to Joe Burke & Finn Butler for proof-reading parts of my thesis.

Thanks to my (both past & currently still present, but soon to be past) office-mates (Finn, Joe, Jon & Tiddy) at the Philosophy Dungeon for conversations so enjoyable that they have kept my mind from going to pieces (for the most part) during the many days (and occasional nights) in the Dungeon. Thanks to Paul Tucek for the countless conversations about philosophy we've shared over the years. Occasionally, it does feel like we have made progress.

In gratitude, I also recognize that the research for this MA-thesis has been supported by the Alan Musgrave Master's Scholarship in Philosophy. As everyone who has ever had the good fortune to almost exclusively dedicate themselves to philosophical research for a whole year will surely attest, philosophy often is uniquely frustrating, unsettling and difficult. It is also uniquely fun, in a way that is hard to describe and communicate to people who have never gotten lost in the vortex of, say, contemporary epistemology. I am deeply thankful to the University of Otago and the Department of Philosophy at Otago for allowing me to throw myself into that vortex.

To finish, I'd like to thank my family, and the one's close to my heart, whose adamant love gives me something to always hang onto.

ABBREVIATIONS

AMS: The **A**rgument from **M**odal **S**ecurity.

EDA: **E**volutionary **D**ebunking **A**rgument.

NNMR: **N**on-**N**aturalist **M**oral **R**ealism.

CONTENTS

0 INTRODUCTION

Genealogical debunking arguments in moral epistemology try to call into doubt that our moral beliefs constitute moral knowledge. They try to do so by attempting to establish the claim that what we know or what we have good reason to believe about the causal origin of our moral beliefs reflects negatively on the epistemic standing of those beliefs. A causal factor that has been the subject of intense debate recently is evolution.

Evolutionary debunking arguments [EDAs] in moral epistemology[1] generally aim to show that what we know or what we have good reason to believe about the impact of evolution on the development of human moral psychology threatens the epistemic standing of our moral beliefs, and therefore we have a *prima facie* reason to withhold judgment in moral matters.[2] Very often though, this claim is qualified so as to apply only to meta-ethical theorists of a certain stripe. The prime target here has been a strong kind of moral realism. In those cases, it is argued that the sceptical conclusion would only follow if we assume strong moral realism. In turn, this is then taken to be a strong reason for rejecting those theories targeted by the respective EDAs. EDAs are thus often construed not as arguments for moral scepticism as such, but as arguments against strong kinds of moral realism. Strictly speaking, what these more restricted arguments try to debunk, is not morality, but a strong moral realism.

In this thesis, I will look specifically at certain replies to EDAs that have been developed on behalf of a strong moral realism. I will argue that certain popular responses to EDAs, which I call the "standard responses" (in virtue of their

[1] Unless indicated otherwise, the label "EDA" will refer to EDAs in *moral epistemology* throughout (as opposed to e.g., EDAs in *religious epistemology*).

[2] I will assume throughout that our moral judgments are correctly conceived of as expressing beliefs (or belief-like states), that e.g., like other beliefs can be properly assessed along epistemic dimensions. This has been disputed by some non-cognitivist meta-ethicists (cf. Ayer 1946 & 1954). An interesting sub-debate concerns the question of whether non-cognitivist meta theories might nevertheless be potentially subject to EDAs, cf. Joyce (2013) and Street (2011), and the reply to Street by Gibbard (2011).

popularity), are ill-suited for neutralizing the epistemic threat that supposedly arises from evolutionary considerations.[3]

Though my arguments may well also apply to other members of the family of standard responses, I will be specifically concerned with the perhaps most prominent members of this family. These are the so called "third-factor" accounts or theories, developed and defended by several philosophers (cf. e.g. Enoch 2010 & 2011; Wielenberg 2010), and the argument from modal security [AMS] developed by Justin Clarke-Doane (2015 & 2016).

There are empirically-minded replies to EDAs, which e.g. argue that what we know or what we have good reason to believe about the impact of evolution on the development of human moral psychology does not support the sceptical conclusion that the debunkers are aiming for, even if we assume a strong moral realism (cf. appendix C).

But the standard responses are importantly distinct from these empirically-minded replies: these standard responses promise that even if the debunkers' empirical story is largely correct, and assuming strong moral realism, the epistemic threat supposedly arising from evolution can be neutralized. So, even if EDAs do not succumb to the empirically-minded responses levelled against them, these responses promise to protect strong moral realism from the charge of leading to moral scepticism. This promise is what makes the standard responses interesting.[4]

I will argue that these responses cannot make good on this promise. Assume strong moral realism is true. And assume a plausible desideratum on any response to EDAs: any satisfying response should support the claim that evolutionary considerations do not show that our moral beliefs are seriously epistemically deficient or that we are seriously epistemically deficient for holding them. In brief: a good response to the

---

[3] This is the same family of responses to EDAs that have recently been called "first-order replies" by Morton (2018: p. 2). Locke (2014: p. 221, p. 227) calls this same family of replies "minimalist". These are replies which crucially assume substantive moral claims. The family of first-order replies/ minimalist replies/ standard responses also includes responses which are somewhat more direct, and certainly less intricate, than the responses I discuss here (cf. Dworkin 1996; Nagel 2012: p. 105; Parfit 2011: Chapter 32, pp. 493f., p. 531).

[4] If, on the other hand, the standard responses would have to rely on the success of the empirically-minded replies to EDAs, this dependence would render them superfluous. This point will come up again a few times.

EDAs should show that evolutionary considerations do not suffice to render us epistemically criticisable for holding or continuing to hold our moral beliefs. I will argue that the considerations offered by the two standard responses are insufficient for satisfying this desideratum.

I should foreclose a potential source of confusion right from the start, concerning both the focus and the methodology of my investigation. In this thesis, I am solely concerned with a limited issue concerning the epistemology of evolutionary debunking:

> Are certain standard responses to EDAs successful in their own right (i.e., without largely depending on the success of other replies)?

The topic I investigate in the four substantive chapters of this thesis is whether we can find an epistemologically respectable epistemic challenge to our moral beliefs that arises from making the discovery that evolution has influenced the moral beliefs of *contemporary humans*, or, more to the point, that evolution has influenced *our* (i.e., *your* and *my*) moral beliefs in some way – an epistemologically respectable challenge that is resistant to the above two standard responses. This way of framing the investigation is not entirely neutral, as it builds on a couple of idealizing assumptions concerning the nature and the scope of evolutionary explanations.[5]

With this point out of the way, here is the plan for the thesis:

Chapter (1): My task in this chapter is twofold. First, I intend to provide useful context on evolutionary debunking and introduce a few key ideas, which are necessary for the presentation and evaluation of the two standard responses. Secondly, I will construct a generic EDA that is based on a clear and reasonably plausible epistemological foundation, and that can serve as a foil for the two standard responses.

Chapter (2): Here I present the two standard responses, the third-factor accounts and the AMS. These two responses offer considerations meant to neutralize the epistemic threat supposedly arising from EDAs. Both responses suppose that the epistemic threat in need of neutralizing is roughly that evolutionary considerations

[5] For a statement of these idealizing assumptions, and for some motivation for making these assumptions in the context of the current project, see appendix C.

might show that our moral beliefs are subject to a problematic kind of epistemic luck. And both responses crucially assume the truth of our common-sense moral beliefs.

Chapter (3): In this chapter, I intend to do three things. I want to elucidate what it is for a consideration to counter-act or neutralize a supposed defeater. Next, I will consider one possible option for how an alleged defeater can be neutralized (the defeater-defeater option). And I will assess the success of the standard responses, if we understand these responses as defeater-defeaters (=conditions or considerations which are meant to reinstate lost justification via counter-acting an existing defeater for a belief). Ultimately, I will draw on recent work by Andrew Moon (2016) to argue that the two standard responses to EDAs are unsuccessful if they are regarded as defeater-defeaters. The reason for this is that it is clearly epistemically wrong to continue to assume the truth of your belief in the presence of a defeater for that very belief.

Chapter (4): In this chapter, I will investigate how third-factor responses and the AMS fare if we understand them as defeater-deflectors (=conditions or considerations which are meant to prevent the occurrence of defeat in the first place). I will argue that even if the standard responses succeed in protecting our moral beliefs' justification, they are nonetheless deeply unsatisfying, as they recommend a kind of doxastic behaviour that is epistemically vicious. Along the way, I hope to show that even getting to this (still deeply unsatisfying) result requires that we buy into a highly controversial epistemological account of epistemic justification and defeat.

I will then summarize my results in the **Conclusion**.

The **Appendices (A-I)** to this thesis contain material that clarifies background assumptions that underlie my reasoning in chapters (1-4), that provides additional detail or lends some further motivation to certain points. Throughout the text, I will refer the interested reader to these appendices, when I lack the space to elaborate on the relevant points in the main body of this thesis or when doing so would side-track us. The material contained in these appendices is (strictly speaking) for the most part dispensable to the main line of argument I develop. Nevertheless, I hope that this material serves an auxiliary function in supplying additional detail, clarification and motivation.

1 EVOLUTIONARY DEBUNKING

1.1 INTRODUCTION

My task in this chapter will be to set the stage for the line of reasoning constructed in the following chapters. In sections (1.2) and (1.3), I will introduce two influential EDAs and a few key ideas, that we need to have a firm grip on to discuss and critically assess the two prominent replies to EDAs, which will be presented in the next chapter. In section (1.4), I construct a generic EDA, which I try to base on a clear and plausible epistemological foundation and which can serve as a foil for the two standard responses.

1.2 TWO EDAS

In this section, I will briefly present two influential EDAs to give us a better initial idea about what evolutionary debunking is about. In section (1.2.1), I will present Sharon Street's Darwinian dilemma. In section (1.2.2), I will present a sketch of Richard Joyce's EDA. Street's and Joyce's EDAs were both instrumental in sparking the recent explosion of interest in evolution in moral epistemology and in shaping the subsequent debate, and so are well worth our attention here as important points of reference for the debate on evolutionary debunking.

1.2.1 STREET'S DARWINIAN DILEMMA

Street (2006) argues that a certain kind of meta-ethical realism is not easily compatible with the theory of evolution. This is due to the Darwinian dilemma Street sets up for this kind of realism. Since Street also argues that anti-realist theories can circumvent the Darwinian dilemma, this seems to make anti-realism the superior option to realism.

The target of Street's argument is a robust kind of meta-ethical realism.[6] Such realist views hold that there are at least some normative/moral facts or truths which hold independently of our actual or hypothetical evaluative attitudes. We might call this

[6] Street's Darwinian dilemma targets not just realism about morality, but rather realism about practical normativity. For the sake of convenience, I have mostly written as if her argument was more narrowly focused, because that won't make a difference in this essay, and it allows for an easier comparison between Street's EDA and other EDAs.

thesis the "stance-independence of morality".[7] The stance-independence of morality entails that moral truths are not constituted or made true by anyone's conative or cognitive stance with respect to them. Proponents of this thesis hold that the truths in the domain of morality are not made true "by virtue of their ratification from within any given actual or hypothetical perspective" (Shafer-Landau 2003: p. 15)

The opening premise of the Darwinian dilemma then goes as follows:

> ***The Darwinian Premise***. Evolutionary factors have had a great (albeit indirect) influence on the content of our evaluative attitudes.

This premise comprises the empirical part of Street's argument.[8]

Street argues that making certain kinds of evaluative judgments tends to result in having higher chances for reproductive success (e.g. "The fact that something would promote one's survival is a reason to do it"), whereas making other kinds of evaluative judgements tends to result in lower chances for reproductive success (e.g. "The fact that something would promote one's survival is a reason against it") (Street 2006: p. 114). Evaluative judgments of the kind that tend to result in higher chances for reproductive success are among the most deeply and widely held judgements we actually hold (ibid.: p. 115). Furthermore, there's a striking continuity between human

---

[7] This is Street's own specific way of drawing that line (2006: pp. 110-111). One should keep in mind that this is not the only possible way of making the distinction between realism and anti-realism, and it is probably not the most common one. A more common way might be via asking the two questions of "Do moral/normative claims (that imply or presuppose the instantiation of a moral property) purport to report facts in light of which they are true or false?" and "Are some moral/normative claims actually true?" (Sayre-McCord 2016). If you answer "Yes" to each of those questions, then you are typically classified as a realist.
As a consequence of Street's somewhat unorthodox taxonomy, some philosophers who regard themselves as realists in some other sense (for instance, because they do hold that moral/normative judgments are both truth-apt and some are actually true) would be counted as anti-realists following Street's taxonomy, and therefore their views are not targeted by Street's argument.
Street has defended her taxonomy by arguing that her way of making the distinction between realism and anti-realism revolves around a crucial point of contention. This is the point of whether or not at least some moral/normative facts or truths hold independently of all our evaluative attitudes, and therefore of whether or not these moral/normative facts and truths are objectively binding (Street 2008: p. 223). Street holds that this point marks out a crucial divide between meta-ethical theorists. The views of those theorists, who assert that some moral/normative truths and facts are objectively binding, are targeted by Street's Darwinian argument.

[8] Street adds the caveat that her conclusion is conditional on her assumptions about evolution being true.

evaluative judgements and the more basic evaluative tendencies of other animals, especially those animals that are most closely related to humans (ibid.: pp. 117-119).

Two complications arise from the opening premise of the argument: (a) Our pre-reflective tendencies to act in certain ways are evolutionarily prior to our making of full-fledged reflective judgements. Moreover, (b) tendencies to make certain evaluative judgments are not genetically hereditary, and therefore this trait cannot evolve through natural selection.[9]

The gist of Street's dealing with (a) and (b) is that the influence of evolution on the content of human evaluative judgments is indirect: Had the general content of our basic evaluative tendencies been very different, then the general content of our full-fledged evaluative judgments would have been very different. Therefore, according to Street there is a loose correspondence between evolutionary impact and human evaluative judgements (ibid.: p. 120). So much for the empirical part of Street's argument.

Given the Darwinian premise, the realist must take a stand on the relation between (i) the evolutionary forces that have had a tremendous influence on the content of our evaluative attitudes, and (ii) the attitude-independent moral/normative truths or facts posited by the realist.

And this leads to the Darwinian dilemma (cf. ibid.: p. 121):

> ***First Horn*** $_{DD}$***.*** The realist denies a relation between the influence of evolutionary forces on our evaluative judgments and moral/normative truths. But this leads to scepticism.
>
> ***Second Horn*** $_{DD}$***.*** The realist asserts the relation, but then she would have to offer an account of how the influencing evolutionary forces and independent

[9] As Mogensen (2016b) points out, the claim that traits need to be genetically hereditary to evolve through natural selection is dubious:
"Selection occurs when the members of a population vary in ways that lead some to survive and reproduce at greater rates than others, provided that those traits which lead to increased reproductive success are reliably transmitted across generations. The exact mechanism by which favoured traits are transmitted is not specified by the definition of natural selection…Thus, traits which are transmitted across generations as a matter of social learning can in principle evolve by selection and constitute adaptations." (p. 1808)
Thanks to James Maclaurin for also raising this point in conversation.

moral/normative truths are linked. But even the best realist account on this link is scientifically flawed.

To the first horn: If the realist denies the relation this leads to scepticism. Why? Well, if the realist denies a relation, then our evaluative judgments (which have been greatly influenced by evolutionary forces) are most likely to be false or even if they were true, they would be true by complete happenstance (ibid.: p. 122). Of course, the realist here is free to assert that indeed our evaluative judgments are true by coincidence. However, Street argues that this is highly implausible, given the range of conceptually possible alternative options. Street compares this with trying to get to Bermuda by jumping in a boat and letting the tide and the wind take you where they may. It is always possible that you will get to Bermuda, but if you do, it will be an astounding accident. In addition, the question then arises, if there is no correlation between the moral/normative truth and our moral/normative judgments, and our judgments just happen to be true, why are we justified in holding them? Justification and knowledge are commonly thought to exclude epistemically lucky coincidences.

If the realist were to reply that we could still get back in touch with moral truths by rational reflection, Street counters this by stating that the success of rational reflection depends on not starting the reflection process with largely false judgments (cf. ibid.: pp. 123-124). Rational reflection works via assessing a given judgment with respect to its consistency with other judgments. If we were to start the reflection process with largely false judgments, then we could only assess false judgments in terms of other false judgments. Then rational reflection is not going to get us closer to the truth. Therefore, if the realist denies the relation between the evolutionary forces that have influenced the content of our evaluative judgments and independent moral truths, then this leads to scepticism.[10]

To the second horn of the dilemma: If the realist asserts the relation, she needs an account on how the evolutionary forces that influenced our evaluative judgments and

[10] As Selim Berker has pointed out, Street does not argue against scepticism, so the realist is free to opt for asserting that there's no relation between our normative judgments and the normative truth and for thereby accepting scepticism (cf. Berker 2014: p.11). However, as Berker himself comments, this is unlikely to strike most realists as an attractive avenue for a rejoinder to the Darwinian dilemma.

the independent moral truths are linked. However, the most plausible account the realist can offer is scientifically flawed. The best account the realist can offer is the tracking account:

> ***The Tracking Account.*** Evolutionary forces have tended to make our evaluative judgments track the attitude-independent evaluative truths or facts because making true evaluative judgments (or proto-versions of these) promoted our ancestors' reproductive success (ibid.: pp. 125-126).

Since the tracking account is intended to be a scientific explanation, it can be assessed by the same scientific criteria (parsimony, clarity, explanatory power) as competing explanations. The trouble for the realist is that with respect to all three criteria, the tracking account fares worse than an alternative account, the so called "adaptive link" account:

> ***The Adaptive-Link Account.*** Evolutionary forces have pushed us towards making certain evaluative judgments because (i) having the basic evaluative tendencies, from which these judgments have developed, made our ancestors more likely to act in accordance with them, and (ii) it promoted reproductive success to act in those ways in our ancestors' circumstances (cf. ibid.: pp. 127-133).

The adaptive link account is more parsimonious because it can explain everything that the tracking account can, without additionally having to posit the existence of independent moral truths. The adaptive link account is clearer because it works without having to rely on independent evaluative truths, and it is highly unclear why exactly it would have been more evolutionary advantageous to track these truths. And the adaptive link account offers more explanatory force because it explains certain

facts about our way of making evaluative judgements that the tracking account can say nothing about.[11, 12]

Since with respect to these three crucial criteria the best account available to the realist is worse than an alternative account, the realist cannot plausibly assert the connection while staying a realist.

Given the Darwinian dilemma, realism about values cannot be easily reconciled with the theory of evolution. What is left to do for Street then is to argue that her own constructivist account can evade the Darwinian dilemma. Now, a complex of various causal forces (one of these forces is evolution) brings about our basic evaluative judgements. Consequentially, Street's constructivist will assert that there is a relation between evolutionary forces and evaluative truths (ibid.: p. 153).

The genesis of the basic forms of our evaluative judgements will be explained via the plausible adaptive link-account (ibid.: p. 153). Evolutionary causes (among other factors playing a role) thus give us a base set of evaluative judgements. We can then go on in refining these basic evaluative judgments by assessing them through rational reflection, testing the consistency of a given judgement with respect to other judgements. The anti-realist (i.e., Street or the Street-style constructivist) now just defines the evaluative truth (or the truth of our evaluative judgements) as being the result of such a process (taking basic evaluative judgements and applying rational

---

[11] For example, it explains why we tend to make certain evaluative judgments (e.g., "That someone is not part of our group is a reason to treat that person worse") that we then would reject on reflection. That explanation would hold that making these judgments promoted ways of acting that were advantageous with respect to survival and reproduction in our ancestors' circumstances. Furthermore, the account explains why we don't make certain evaluative judgements (e.g., "That something decreases one's chances of survival is a reason to do it").

[12] Her comparison between the tracking and the adaptive link-accounts raises a host of questions that Street unfortunately does not address. For instance, are there evaluative judgements we regard as true even on reflection, yet they wouldn't obviously have been evolutionarily advantageous? In addition, are there evaluative judgements we regard as false on reflection, yet they would plausibly have been evolutionarily advantageous? This boils down to the question whether we can give an evolutionary explanation of the content of conventional morality. Second, why can't the tracking account say e.g., that we don't make the judgement that plants are more valuable than human beings because making false evaluative judgements was evolutionarily disadvantageous (in the same way that making true evaluative judgements was advantageous)? Thanks to Robin McKenna for raising these points.

reflection to them) (ibid.: pp. 153-154). In conclusion, since anti-realism fits better with the theory of evolution than realism, that speaks in favour of anti-realism.

1.2.2 JOYCE ON THE EVOLUTION OF MORALITY

Whereas Street thinks, moral scepticism only follows if we assume a strong form of moral realism, Joyce (2016a: pp. 145-146) doubts that meta-ethical presuppositions that turn on endorsing or rejecting the stance-independence of morality can make a difference to the debunking capacities of evolutionary explanations vis-à-vis our moral beliefs.[13] His own EDA is not an argument that primarily targets any particular meta-ethical view such as a robust version of realism. Rather, Joyce's EDA is an argument for moral scepticism, i.e. the claim that all our moral beliefs are unjustified (ibid.: p. 146).

Furthermore, while Street proposes that evolutionary forces have heavily influenced the content of our evaluative beliefs, Joyce (2006) advances an evolutionary explanation for our tendency to make moral judgements of any sort. He supports an account of the evolutionary origin of the human moral faculty termed "moral nativism". According to moral nativism, humans possess a "specialized innate mechanism (or series of mechanisms)" that "comes prepared to categorize the world in morally normative terms" (Joyce 2006: pp. 180-181). In other words, we humans have an innate tendency to make moral judgments.[14]

The empirical component of Joyce's argument is the claim that this innate moral sense is a feature shared by all normal humans combined with an evolutionary explanation of why humans have this innate moral sense.

This evolutionary explanation holds that moral thinking was selectively advantageous to our ancestors because it reliably enhanced cooperative behaviour in our ancestors'

---

[13] This does not mean that other meta-ethical presuppositions might not be able to make a difference in this respect. Consider for instance the following line of thought: moral judgments can only be evaluated in epistemic terms if they are fit to express beliefs. A non-cognitivist meta-ethicist, who rejects this assumption, therefore seemingly has an easy way of sidestepping EDAs. An interesting debate has developed concerning the question of whether non-cognitivist views are potentially also subject to EDAs (Gibbard 2011; Joyce 2013; Street 2011).

[14] For an up-dated and detailed discussion of the troubles with defining "moral nativism" and defending a version of the view, cf. Joyce (2016c: pp. 122-141) and Sripada (2008).

circumstances.[15] Our ancestors' capacity for moral thought allowed them to conceive of certain behaviours (e.g., helping someone else) as being required in such a strong sense that this requirement could not be trumped by other interests (which e.g., called for non-cooperation) (Joyce 2006: pp. 60-63). Jessica Isserow sums up the bottom line of Joyce's empirical hypothesis nicely:

> Moral faculties therefore earned their evolutionary keep by enabling individuals to block competing interests that would interfere with prosocial motivation from the deliberative sphere—they functioned as devices of "personal commitment"…Given that prosocial behaviour is often costly, moral faculties also came to serve as signs of interpersonal commitment; they offered early humans a means by which to convincingly signal their prosocial dispositions to others. (2018: p. 5)

A noteworthy feature of this explanation is that it neither explicitly appeals to nor requires the assumption that there actually exist any moral truths or facts.[16] As Joyce writes, "…whether we assume that the concepts 'right' and 'wrong' succeed in denoting properties in the world…the plausibility of the hypothesis on how moral judgments evolved remains unaffected" (2006: p. 183). We might call this the "INDEPENDENCE THESIS":

> ***INDEPENDENCE THESIS***. The plausibility of the evolutionary explanation of why we have a propensity for making moral judgments is independent of the assumption that any of our ancestors' moral judgments were true.

As Joyce (2013) notes, the INDEPENDENCE THESIS does not hold across the board for all kinds of evolutionary explanations, i.e. sometimes the best evolutionary

---

[15] Furthermore, Joyce writes that apart from their capacity for moralized thought, humans also have prosocial inclinations that dispose us to cooperate with others (2006: pp. 47-51). But these inclinations fall short of reliably assuring cooperation, which is the reason for the need to develop a capacity for moral thought that goes "above and beyond a suite of emotional dispositions" (Isserow 2018: p. 5).

[16] But what if you endorse a meta-ethical theory that holds that the moral facts are reducible to whatever improves social cooperation (or something in that vicinity)? Wouldn't Joyce's claim be false on the assumption that a version of moral naturalism is true? This is correct, and that is why a considerable part of Ch. 6 of Joyce (2006) is dedicated to arguing against naturalism. For more on moral naturalism and EDAs, cf. also Barkhausen (2016) and Locke (2014: pp. 225-227).

explanation for a certain trait will classify this trait as truth-tracking. Take e.g., the case of humans possibly having an adaptive mechanism for distinguishing faces from other visual stimuli. The best evolutionary explanation for the face-identifying faculty classifies it as truth-tracking (i.e., a mechanism of this sort will not be fitness-enhancing if it is does not produce accurate beliefs), while Joyce contends that the best explanation for the moral faculty does not (2013: p. 353). In the case of the face-identifying faculty, the reason why having this trait might increase the fitness of its bearers directly refers to the fact that this faculty is truth tracking. Plausibly, the ability to distinguish human faces from other visual stimuli confers a selective advantage on its bearers (if it does) only if it tracks the relevant truths about human faces. In the case of our moral faculty on the other hand, Joyce holds that our ancestors' propensity for making moral judgments plausibly conferred a selective advantage upon them independently of whether those judgments track the moral truths.

Herein Joyce echoes Michael Ruse (1986) who argued that morality serves its adaptive function by strengthening our motivation for social cooperation through seeming to be imbued with a kind of external prescriptivity (1986: p. 103). That is the adaptive importance of the objectivity with which moral prescriptions are infused. However, this objectivity is an adaptive illusion (ibid.: p. 253).

Ruse argues for this claim with an implicit appeal to parsimony: once we have explained why morality seems to be objective, there is simply no further call for explaining it in terms of positing objective moral facts (Joyce 2013: p. 356). Apparently, the evolutionary explanation works perfectly fine without our positing the existence of these facts. Therefore, the INDEPENDENCE THESIS seems to hold for the evolutionary explanation of our moral beliefs.

This has epistemic consequences. Plausibly, a belief's justification depends on it standing in a certain relation to the fact it represents. If evidence supports moral nativism, then this seems to be a confirmation that our moral beliefs have their origin in a process that is not set up to track the truth of the beliefs it produces.

Joyce claims that this undermines the epistemic standing of our moral beliefs, as this discovery is relevantly like the discovery that a few years ago you have taken a pill, which made you form lots of beliefs about Napoleon, and without having taken the

pill you would not have formed any beliefs concerning Napoleon at all (2006: p. 181). Upon making this discovery, your beliefs about Napoleon lose all the justification they previously might have enjoyed. After all, you lack any reason to think that this belief-forming method relates your Napoleon-beliefs to their subject matter in an appropriate and reliable way. And this makes that method untrustworthy.

Joyce thinks that similarly, upon discovering the evolutionary origin of our moral thinking, our moral beliefs lose all the justification that our beliefs might have enjoyed before (2013: p. 353). Given the evolutionary explanation of our moral beliefs, we too lack any reason to think that our moral belief-forming methods relate our moral beliefs to their subject matter in a reliable way. Therefore, it seems that this shows our beliefs to be the product of an untrustworthy method. In this way, Joyce aims to establish a kind of moral scepticism, by arguing that given moral nativism, our moral beliefs are unjustified.[17]

1.3 KEY IDEAS

In this section, I will introduce and explain several concepts, which are important to my project. In section (1.3.1), I will talk about the kind of evolutionary explanation involved in most EDAs. In section (1.3.2), I will talk about defeating epistemic justification. In section (1.3.3), I will present the favourite target of many EDAs, non-sceptical, non-naturalist moral realism. In sections (1.3.4) -(1.3.5) I will then present ideas which are integral to understand the two standard responses (especially Clarke-Doane's modal security-argument).

[17] Joyce originally labelled the result of his EDA (combined with arguments against moral naturalism that can be found in the last chapter of Joyce (2006)) an error theoretic conclusion (2006: p. 223). This was motivated by his view that an error theory of morality might be understood as the disjunction of two accounts. That means an error theory denotes either the view that all moral judgements are false or the view that all moral judgements are unjustified. In a more recent publication (2013: pp. 354-355), Joyce recants this position for the following reason. Suppose moral scepticism is true, then all moral beliefs are unjustified. This claim is consistent with our moral judgments being true. Thus, moral scepticism is consistent with a (sceptical) version of moral realism. Hence, moral scepticism does not suffice to establish the truth of an error theoretic conclusion. In turn, Joyce suggests a new error theoretic argument that turns on the explanatory impotence of objective moral facts (2013: pp. 355-363).

1.3.1 EVOLUTIONARY FORCES AS TRUTH-IRRELEVANT INFLUENCES

EDAs hold that evolutionary explanations challenge moral knowledge (or they challenge moral knowledge given that we assume a certain meta-ethical view). Subsequently, a good point with which to start the discussion is: what kind of evolutionary explanation is involved in those arguments?

The bulk of contemporary EDAs have focused on explanations in terms of natural selection: i.e., beneficial effects on reproductive fitness explain the prevalence or existence of a certain trait (Mogensen 2014: p. 10; the EDAs due to Joyce and Street presented above are prominent examples here).[18] In biology, this kind of explanation is called a "functional explanation". This is an explanation of why certain organisms have certain traits rather than some conceivable alternatives by appealing to the reproductive advantages of having these traits rather than alternatives for those organisms' ancestors (Mogensen 2014: p. 10). For example, in explaining why certain birds display migratory behaviour we might cite the ways in which spending different seasons in different locations was advantageous to the bird's ancestors.[19]

Now, many EDAs appeal to functional explanations of our moral beliefs for the following reason: as Mogensen puts it, they are "narrowly focused on the issue of functional truth-irrelevance" (Mogensen 2014: p. 19). In other words, many debunkers argue that evolutionary explanations are debunking because they provide evidence for the claim that our moral belief forming mechanisms are not connected to the moral truths in the right kind of way, since (i) our moral belief forming mechanisms were selected for on account of their tendency to produce reproductive fitness-enhancing beliefs (or belief-like states). And (ii), these advantages in terms of reproductive fitness are *explanatorily independent* of the accuracy or truth-conduciveness of these beliefs.

The general idea behind these prominent EDAs thus is that if the origin of our moral beliefs is explained in a way that is *truth-irrelevant*, then this should significantly reduce

---

[18] The notable exception here is Andreas Mogensen's EDA (2014; 2016b; 2017). His argument relies on an explanation of morality in terms of our evolutionary inheritance, a *phylogenetic* explanation, and not on the viability of an explanation of our moral beliefs in terms of natural selection.

[19] I take the example from Levy & Levy (2016: p. 2).

our confidence that these moral beliefs are true. These EDAs are therefore driven by the (apparent) insight that an evolutionary explanation in terms of natural selection threatens the prior justification our moral beliefs might have enjoyed by showing how our moral belief forming mechanisms have been shaped by forces that have not primed these mechanisms to produce accurate beliefs.

1.3.2 DEFEATING JUSTIFICATION

EDAs try to establish the point that the evolutionary explanation of our moral beliefs reflects negatively on the epistemic *justification* for those beliefs. It locates the epistemological threat of evolutionary considerations specifically with the notion of epistemic justification. I now want to shed some light on how evolutionary considerations are meant to take away the justification for our moral beliefs. A few brief remarks are in order here.

First, assume a belief of yours constitutes knowledge. Then we remove epistemic justification from this belief. Removing justification is sufficient for making it the case that your belief no longer constitutes knowledge. This is the reason why I am interested in epistemic justification in this thesis. So, I am interested in epistemic justification as a property of *beliefs* – not of *believers* or of *propositions* (cf. Littlejohn 2012: p. 5). Epistemic justification, as I am interested in it here, is about whether a *belief* is justifiably held. And this is epistemologically interesting, because I assume that justification *in this sense* is necessary for knowledge.[20]

Secondly, it should be emphasized that justification is not the only focus of epistemic evaluation, e.g., we can also assess an agent's epistemic standing directly in terms of knowledge or in terms of the agent exhibiting certain epistemic virtues or vices in her doxastic behaviour or in terms of her understanding etc. This point is important because appreciating this lets us see that in principle at least, a debunking argument

[20] So, I will use the term "justification" to refer to a property that is possessed strictly speaking not by *believers*, but by *beliefs*. On this way of using the term, there is only a derivative sense in which believers are justified, i.e. in virtue of believing justifiably (=holding a belief that is justified). To avoid baroque syntax, in a couple of instances I will write a bit more loosely, but nevertheless, the property of justification I am interested in throughout is always the one that is in the first instance a property of beliefs.

could also zoom in on an area of epistemic evaluation different from epistemic justification.

Thirdly, assume that epistemic justification is tied up with the normative and deontic notions of what we ought to believe and what we are permitted to believe: Being justified in believing that p is a necessary condition for being permitted to believe that p and for it being true that you ought to believe that p (Mogensen 2014: p. 6). Therefore, if you are not justified in believing that p, you are not permitted to believe that p, and it is not the case that you ought to believe that p. Furthermore, on the assumption that if you are not permitted to believe that p, then you ought not to believe that p, it follows then that if you are not justified in believing that p, you are not permitted to believe that p, and if you are not permitted to believe that p, then you ought not to believe that p.

Why is it the case that the debunking potential of evolutionary considerations specifically has enjoyed so much attention in recent philosophical discussion? It seems plausible, that this is due to a "felt possibility that we may be forced to make substantial revisions in our moral outlook in light of the new discoveries about the evolutionary origins of our moral beliefs" (Mogensen 2014: p. 6). And as Mogensen points out in the same passage, since epistemic justification is tied up with what we ought to believe and what we are permitted to believe, understanding EDAs in terms of defeating justification accounts nicely for this.

The point of EDAs as they are usually understood is then to establish the claim that we are not justified in holding our moral beliefs and are therefore not permitted to believe them and ought not to believe them (or they try to establish a qualified version of this claim that is conditional on the truth of the meta-ethical view targeted).

We can understand the epistemological threat posed by evolution according to the EDAs in terms of *epistemic defeat* of a belief's justification. What is epistemic defeat, and what are defeaters?

Generally, the *defeasibility* of a belief refers to the belief's liability to lose some or all of its putative positive epistemic status, or to having this status downgraded in some particular way (Sudduth 2008). A belief that p can be *defeasibly justified* in the sense that

this belief enjoys justification (or some level of justification) in the absence of information that counteracts the reasons for believing that p.

A *defeater* is now roughly a condition or consideration that gives us a defeasible reason to suspend belief or reduce our confidence in a proposition that we would otherwise have been justified in believing (Mogensen 2014: p. 7). Since defeaters are defeasible reasons to withhold belief, they can themselves be defeated by conditions or considerations referred to as *defeater-defeaters*. These might themselves be defeated by *defeater-defeater-defeaters*, and so on.

Usually it is held that there are at least two distinct classes of defeaters: rebutting or overriding defeaters and undercutting or undermining defeaters.

A *rebutting* or *overriding* defeater for your belief that p is a reason for believing the negation of p, or for believing a proposition that is incompatible with p being true (Pollock 1986: p. 38).

An *undermining* or *undercutting* defeater for your belief is instead a reason that attacks the connection between your belief p and your grounds for believing that p, it is e.g., a reason to doubt the trustworthiness of your grounds for belief or the method by which it was produced.[21]

From what I have said so far, it should be clear that EDAs do not supply us with a rebutting defeater: They do not give us a direct reason to believe that any of our moral beliefs are false. For all these arguments tell us, they all could be true (as they stand, the EDAs presented in section (1.2) above are perfectly consistent with the truth of our moral beliefs).[22, 23]

---

[21] There are also hybrid-defeaters, i.e. defeaters that are both undermining and rebutting. In some cases, it may be that a rebutting defeater is also undermining. If a certain belief-forming process of mine gives me a wrong result, if I have no independent reason to think that this was due to a random error and this process does not have a sufficient track record of success, then this seems to give me (some) evidence, that the process is systematically biased towards error, i.e. I receive an undermining defeater (Mogensen 2014: p. 7).

[22] This point is important to emphasize, as EDAs would appear to be instances of a genetic fallacy (=inferring the falsity of a claim from its origin), if one were to hold that they show our moral beliefs to be false.

[23] In the context of the debate on peer disagreement, it is often assumed that discovering that somebody (who does not seem to be irrational or who is even on par with you epistemically speaking) disagrees with you can give you a specific kind additional evidence speaking against your belief, i.e. "higher order evidence". Such evidence works by inducing

Rather, the debunkers present arguments that turn on the evolutionary origin of our moral beliefs to generate an undermining defeater (this is explicit in Joyce 2013; Locke 2014; Lutz 2017; Mogensen 2014, 2016a & 2017, it is also a plausible interpretation of Street 2006).

There is also another important distinction between different kinds of defeaters to keep in mind: On the first account, we must distinguish between doxastic or psychological defeaters on the one hand, and normative defeaters on the other. I take this way of drawing the distinction from Jennifer Lackey (cf. e.g., 2014: p. 305).

Doxastic or *psychological defeaters* are doubts or beliefs that the subject S has, which indicate that S's belief that p is either false or formed or sustained in an epistemically untrustworthy fashion. As Lackey writes these beliefs or doubts can be defeating regardless of their own truth or epistemic status (ibid.). The main worry concerning doxastic defeaters is that one can unjustifiably/ irrationally or epistemically irresponsibly believe or doubt that one's belief that p is epistemically inappropriate. Can an unjustified/ irrational or irresponsible belief or doubt take away your justification for believing that p?

Here's a simple argument that seems to support the idea that an unjustified or irrational belief cannot be a defeater. Assume the following is true: S's belief that p at point in time t is justified if and only if S's belief that p at t is supported by S's evidence.[24] Call this view "Simple Evidentialism". Next, assume that necessarily, an unjustified belief of S cannot be part of S's evidence. This is not completely uncontroversial, but let's grant the assumption for the purposes of this paragraph. Then, it seems, it is impossible for an unjustified belief of S to add to or take away

---

doubts that your belief is the result of a flawed process (cf. Christensen 2010). So, your belief may be defeated by higher order evidence (higher order defeat is also not limited to cases of peer disagreement, cf. ibid.). Some philosophers, like Christensen (2010), also think that "higher order defeat" is importantly different from undercutting defeat. Unfortunately, for reasons of space, in this thesis it will be impossible to engage with growing and complex literature on this topic (cf. Horowitz 2014 for a partial overview). In what follows, I will assume that such defeat can be thought of as belonging to the class of undermining or undercutting defeat.

[24] This is the core commitment of a family of popular views called "evidentialism" (cf. Conee & Feldman 2004). Alexander (ibid.: pp. 906-908) also argues that evidentialism, despite initial appearances, does not really support the conclusion that unjustified beliefs cannot constitute defeaters.

from the justification of S's belief that p at t. This simple argument seems initially attractive, but things become more complicated, when we consider the following enticing thought, that clashes with this argument. Many epistemologists think that justification has a "perspectival dimension", i.e. a belief is justified only if it is permissible from one's perspective.[25] And what is permissible from one's perspective is plausibly constrained by the beliefs one holds – even by irrational or unjustified beliefs one holds. As Alexander writes:

> Unjustified beliefs may not be part of one's evidence. But from one's perspective an unjustified belief may *seem* to be part of one's evidence in which case it would be irrational for one to treat it otherwise. (Alexander 2017: p. 900)

This thought strikes me as compelling enough to reject the above simple argument. Denying the perspectival dimension of epistemic justification seems to be costly. The idea that justification seems to be constrained by the believer's perspective will also be important in section (1.4.1) below.[26]

A *normative defeater* is a doubt or belief, that S ought to have and that indicates that S's belief that p is either false or formed or sustained in an epistemically illicit fashion. Defeaters in this sense function by virtue of being doubts or beliefs that S should

---

[25] I take that term from Alexander (2017: p. 908).

[26] In an influential treatment of defeaters, Michael Bergmann (2006: Ch. 6) defends the point that all believed defeaters are actual defeaters since it is a necessary condition on S's belief that p to be justified that S does not take her belief that p to be epistemically inappropriate. Bergmann's (2006: pp. 165-169) response to the above worry is that it runs together two separate questions: on the one hand there is the (i) question of how we ought to change our beliefs from the vantage point of ideal rationality, and on the other hand there is the (ii) question of what doxastic response is appropriate given that we hold certain beliefs. It may be that it is irrational for you to have a certain belief from the vantage point of ideal rationality, but given that you hold this belief, it might nonetheless be epistemically inappropriate for you to have some other belief. Bergmann (ibid.: p. 166ff.) also argues that generally, there is no reason to believe that only justified beliefs can be defeaters via attacking the alleged analogy between doxastic attitudes that can transmit or confer justification onto a belief and doxastic attitudes that can take away justification. Here Bergmann argues that we "have good reason to think that the way justification gets produced via inference is importantly different from the way justification is defeated via beliefs that are defeaters" (ibid.: p. 167).
Alexander (2017) criticizes Bergmann's arguments, but ultimately agrees that beliefs that "lack any positive epistemic merit can be defeaters…[c]onsequently, a belief can be a defeater even if it is neither supported by one's evidence nor is the result of a reliable process – nor possesses any other positive epistemic merit" (p. 911).

have (whether or not S does have them) given the evidence available to S at that time (ibid.).

The epistemic danger to your beliefs that arises from defeaters can be met in the following two ways (cf. Moon 2016: p. 5): One can offer *defeater-defeaters*, i.e. one can offer reasons reinstating our provisionally lost justification by countering defeasibly defeating considerations. That means to have a defeater-defeater for the defeater D, D must already be a defeater, which then gets defeated in turn. Examples for defeater-defeaters include cases where I first get information that constitutes a defeasible defeater for my belief that p, but then I discover that this defeasible defeater is misleading, which in turn defeats my original defeater (cf. Moon 2016: p. 5).

Here is an amended example for a defeater-defeater taken from Moon (2016: p. 6):

> ***Defeater-Defeater Case***: You ingest a pill called XX. You go outside to the driveway, and you see a red car parked there. Plausibly, your perceptual belief that there is a red car in front of you is defeasibly justified. But then you learn that XX makes 95% of those who ingest it hallucinate red cars even when there are no red cars in front of them. Here, you have a defeater for your belief that there is a red car in front of you. Two hours after you have taken the pill, a scientist, whom you know to be trustworthy, informs you that you are one of the 5% who is immune to the drug. The justification for your perceptual belief that there is a red car parked in the driveway is reinstated.

Alternatively, you can offer *defeater-deflectors* to protect your beliefs against defeat, i.e. reasons that are supposed to prevent supposedly defeating considerations from providing defeaters for our beliefs in the first place. A successful defeater-deflector prevents D from being a defeater in the first place. Examples here include cases where I already have or receive information $INFO_1$ at time t_1 that keeps information $INFO_2$ received at a later time t_2 from constituting a defeater for my belief that p at t_2, where, if I had not received $INFO_1$ beforehand, $INFO_2$ would have constituted a defeater for my belief that p at t_2.

Here is again an amended example from Moon (ibid.):

> ***Defeater-Deflector Case***: A scientist whom you know to be trustworthy tells you that you are immune to the effects of the pill XX. You proceed to ingest XX. You

> walk outside and see a red car in the driveway. You form the defeasibly justified belief that there is a red car in front of you. You then learn that XX makes 95% of those who ingest it see red cars even when there are no red cars in front of them. However, it seems that receiving this information now does not reflect negatively (epistemically speaking) on the justification for your belief.

In the Defeater-Deflector-Case, receiving the information that I have ingested a pill never impinges on the justification for my belief. In the Defeater-Defeater Case on the other hand in the time after you learn about the effects of XX, but before you have been told by the scientist that you are immune, the justification for your belief is gone. In that case, but not in the Defeater-Deflector-Case, there is a time where you are not permitted and where you ought not to believe that there is a red car in the driveway.[27]

Epistemic defeat is the most important notion in this thesis, and so it will accompany (or, depending on your attitude towards epistemology, haunt) us throughout. But for now, I hope what I have said so far suffices to give us a good initial grip on this notion.

1.3.3 NON-SCEPTICAL, NON-NATURALIST MORAL REALISM

It is often supposed, that EDAs are especially or even exclusively problematic for the epistemological plausibility of a certain meta-ethical position: non-naturalist moral realism [NNMR].[28] This assumption is questioned and explicitly rejected by some debunkers (cf. e.g., Barkhausen 2016; Locke 2014), but for the purposes of streamlining my presentation, there is no harm in supposing that if any view is in the cross-hairs of EDAs, it is NNMR, while leaving the question of whether debunking worries also apply more generally (e.g., to all non-error theoretic meta-ethical theories) aside. This assumption is justified by the focus of this thesis on responses, which (for the most part), have been brought forward by non-naturalist moral realists to defend their view against EDAs. NNMR is the view that is defended against EDAs by the standard responses on which I will focus in the chapters to come. So, let's assume

---

[27] For a brief explanation of why the present framework of epistemic defeat is not begging the question against epistemic externalism or internalism, please see appendix A. Cf. also appendix E.

[28] Cf. Crowe (2016); Lutz (2017).

that the debunkers aim to show that moral knowledge is somehow out of reach *if NNMR is true* – which in turn would give us a strong incentive to reject NNMR.

Now, NNMR is comprised of several positions in the fields of semantics, moral psychology, metaphysics and epistemology. Here is a rough and ready characterization of the view, which I have taken from Schechter (2018) (although I have tweaked it a bit):[29]

(a). **Semantic Factualism**. At least some moral sentences express propositions and purport to represent moral facts, and they are therefore meaningful and truth-apt.

(b). **Cognitivism**. Moral judgments express beliefs or are belief-like states that aim to represent how the world is in a certain respect, and consequently they can be assessed on whether they represent this aspect of the world accurately or not, i.e. on whether they are true or false.

(c). **Non-error theory**. Some atomic moral sentences (i.e., sentences that suppose or imply the instantiation of a moral property) are actually true.[30]

(d). **Independence**. The explanatorily basic moral facts do not depend on us, they are "stance-independent". Facts about our minds, language or social practices do not constitutively explain the fundamental moral facts. That means that the fundamental moral facts obtain "independently of any preferred perspective, in the sense that the moral standards that fix the moral facts [i.e., the explanatorily basic moral facts] are not made true by virtue of their ratification from within any given actual or hypothetical perspective" (Shafer-Landau 2003: p. 15).[31]

---

[29] For example, conditions (a) and (b) above form only one condition in Schechter (2018). I split them up to account for the fact that some meta-ethicists could endorse either (a) or (b), but not both (cf. e.g., versions of meta-ethical fictionalism).

[30] Roughly, an error theory of morality amounts to the conjunction of the following positions: (i) Semantic factualism, (ii) Cognitivism, and (iii) all atomic moral sentences are false. In other words, moral judgments express genuine beliefs and they constitute genuine assertions, and each moral judgment that implies or presupposes the instantiation of a moral property (e.g., "Breaking your promise to Peter is morally wrong") is false.

[31] Given the above definition of Independence, it is conceivable that both the actual moral deliberators and hypothetical idealized moral deliberators could be wrong about what the explanatorily basic moral facts are (although NNMR will of course hold that we are actually not wrong about all of the basic moral facts). In this sense, the explanatorily basic moral facts transcend our (actual and hypothetical) perspective on them. But NNMR is held to be

(e). **Non-plenitude.** Of the many possible coherent practices of moral assessment, only a few are correct. Not all practices are on par.

(f). **Non-Naturalism.** There is a distinct class of moral facts and properties, and these are non-natural and non-supernatural facts and properties.[32] Importantly, these distinct moral facts and properties are causally inert (which distinguishes them from natural facts and properties).

(g). **Epistemic Success**. Many of our moral beliefs are successful in constituting moral knowledge. That means (i) by-and-large, the moral claims we believe upon reflection and discussion are epistemically justified, and (ii) by-and-large, the moral claims we believe upon reflection and discussion are true, or at least, we do significantly better than chance would predict.[33]

In section (1.4), when I formulate a generic EDA, I will suppose that NNMR is the view targeted by this generic argument.

1.3.4 METAPHYSICAL NECESSITY OF BASIC MORAL TRUTHS

Third-factor responses assume that basic moral truths are necessary in a fairly strong sense (Enoch 2011: p. 172). The AMS rests on the assumption that the basic moral truths (if they exist) are metaphysically necessarily true. In this section, I want to elucidate this assumption.

Here, the basic moral truths or facts are true propositions that state the conditions under which a moral property is instantiated. The basic moral truths take the form:

If x has property N, then x has property M.[34]

---

consistent with a weak form of mind dependence that basically amounts to the insight that there would e.g. be no obligations if there were no rational agents to be obligated (Shafer-Landau 2003: p. 15).

[32] Following Miller (2003: p. 4), I take natural properties and facts to be those which are either causal or detectable by the senses, which are studied by the typical "natural" sciences or by psychology. Supernatural facts and properties are understood as non-natural facts and properties that imply or suppose the activity of the kind of non-material beings typically invoked by religion, mythology or mysticism like e.g. God or gods, ghosts, spiritual forces or some such.

[33] Importantly, (g) also seems to presuppose that even before reflection and discussion take place, we are at least not epistemically hopeless in our intuitive moral beliefs. If all our initial moral beliefs were epistemically unjustified, it is hard to see how reflection and discussion could remedy this situation. This harkens back to a point by Street from section (1.2.1).

[34] Cf. also Shafer-Landau (2003: p. 15).

Where N is some (possibly very complex) non-moral property (that gives an exhaustive description in non-moral terms),[35] M is a moral property and where x is some entity fit for being ascribed a moral property (i.e., usually some action or person). Basic moral beliefs have as their content propositions of this form.

Therefore, the AMS involves the claim:

> **M-NECESSITY**. It is metaphysically necessary that for all x: if x has property N, then x has property M.

What's the rationale for positing this metaphysical necessity-claim? NNMR holds that moral properties are not natural properties, i.e. they are neither identical to any natural property, nor reducible or explainable in terms of natural properties. But now consider the following highly intuitive thought: No two actions can differ in their moral properties without differing in their natural properties (and conversely, two actions with identical natural properties also have the same moral properties). For example, if I lie to my friend in scenario A, and if I lie to my friend in scenario B, and everything in scenario A is exactly like it is in scenario B when it comes to all of the natural properties (including my lying to that friend), then it can't be that the two cases differ morally, it can't be e.g., that it is morally bad or impermissible to lie in A, but morally good or permissible to lie in B.

To account for this highly intuitive thought, proponents of NNMR (but also meta-ethicists more generally) typically posit that the moral supervenes on the natural, i.e. no two things or states can differ with regard to their moral properties without also differing with regard to their natural properties. A strong supervenience-relation holds that necessarily whenever something has a moral property M it has some (possibly very complex) non-moral property N that necessitates M, in the sense that necessarily anything that is exactly alike it with respect to the property N also has the moral property M. In other words:

> **STRONG SUPERVENIENCE**. As a matter of conceptual necessity, when something has any moral property M, it has some (possibly very complex) non-

[35] So e.g., if x is some act, then N gives a complete specification of all its non-moral features, such as the acts "intrinsic nature, its causes and effects, the intentions with which it was done, and so on, insofar as these can be specified in wholly non-normative terms" (Rosen Msc.: p. 1).

> moral property such that, as a matter of metaphysical necessity, anything that is exactly alike it with respect to the non-moral property also has the moral property.[36]

This supervenience-relation between the moral and the natural invokes two necessity claims. The outermost necessity claim is most often interpreted as conceptual (i.e., being necessary in the very same sense in which all bachelors are unmarried), while the innermost necessity claim is often interpreted as metaphysical (i.e., being necessary in the very same sense that the claim "Atoms of gold contain 79 protons" is necessary) (cf. McPherson 2015; Väyrynen forthcoming: p. 5).

The strong supervenience-relation between the moral and the natural entails the above metaphysical necessity-claim that anything that has some (possibly very complex) non-moral property has that moral property as well (Väyrynen forthcoming: p. 5). In other words, M-Necessity, which tells us that as a matter of metaphysical necessity, if something has a non-moral property N, then it has a moral property M, is entailed by the claim that as a matter of conceptual necessity, when something has any moral property M it has some (possibly very complex) non-moral property N, such that as a matter of metaphysical necessity if it has N, then it has M.

Therefore, it seems that M-NECESSITY is a consequence of STRONG SUPERVENIENCE. M-NECESSITY is then well motivated if STRONG SUPERVENIENCE is.

As far as I can see, the best motivation for STRONG SUPERVENIENCE is provided by the fact it gives us a general account about the relation between two families of properties (moral and non-moral) that rules out the possibility mentioned above that there could be two cases, which are identical with respect to their non-moral properties, but which differ with respect to their moral properties.[37]

---

[36] I am glossing over complications pointed to by McPherson (2012: pp. 210-220).

[37] In a similar vein, Tristram McPherson (2012: pp. 211ff.) also considers the question of how a strong supervenience-thesis is motivated, and he takes it that it is best motivated through a two-part process:

(i) We start by considering various particular cases which are ruled out by specific instances of a strong supervenience-thesis (examples for cases where two scenarios have all the same non-moral properties, but which differ in their moral properties).

1.3.5 MODAL CONDITIONS ON KNOWLEDGE

To understand the AMS, we also need to introduce two prominent modal conditions on knowledge: safety and sensitivity. Many epistemologists think that either safety or sensitivity are necessary for knowledge.[38] Before talking about these two conditions, let me provide some context.

Modal conditions necessary for knowledge are introduced by epistemologists to help answer the following question:

> What relation must hold between a fact that p and the belief that p in order for the belief that p to amount to knowledge?

It is widely agreed that knowledge is incompatible with accidentally true belief, cases where your belief is related to the corresponding fact only by accident. This has often been taken to be the upshot of Gettier-cases, brought forward originally as counter-examples to the "justified true belief"-analysis of knowledge (cf. Gettier 1963).

Here's an example for such a case (Rabinowitz 2011: section 1; Russell 1948): Suppose Bertrand truly believes that it's noon as a result of looking at his clock that correctly reads noon. However, what Bertrand does not know is that his clock broke down exactly twelve hours prior. Even though Bertrand seems to have good reasons to believe that it's noon and his belief is true, Bertrand clearly does not know it's noon since he is lucky that his belief is true.

Modal conditions are often used in epistemology to cash out the intuitive idea that knowledge requires our beliefs to be not merely correct, but to stably track the truth even if your circumstances were to slightly change (Ishikawa & Steup 2017: section

---

(ii) We notice that our view about the specific cases do not seem to rest on idiosyncrasies of those cases. This encourages the inductive thought that it is impossible for there to be an example where it is not the case that the moral supervenes on the natural. Furthermore, a general strong supervenience-thesis is explanatorily attractive: it is simpler than having to posit and explain a huge raft of specific necessary connections, and it suggests the hope of explaining it in terms of quite general features of moral properties.

Thus, McPherson argues that the commitment to a general strong supervenience-thesis is best seen as being (epistemically) grounded in an elegant and seemingly unproblematic generalization from uncontroversial cases. Cf. also Rosen (Msc. pp. 1-2).

[38] That is to say, these conditions are often proposed as necessary conditions for knowledge, but they are usually not taken to be sufficient.

5). The hope is that by applying either of these modal conditions, we can filter out cases that involve epistemic luck of the kind that is incompatible with knowledge. So, here are the two conditions stated:

> ***Sensitivity$_3$***. S's belief that p formed via or based on M is sensitive if and only if: in nearby possible worlds where p is false, S would not believe that p via or based on M.
>
> ***Safety$_3$***. S's belief that p formed via or based on M is safe if and only if: in nearby possible worlds where S believes that p via or based on M, p is true.[39]

Sensitivity is intended to capture an important feature of knowledge, namely the ability to tell the difference between when a proposition is true and when it is not (Becker 2012: p. 82). Safety is meant to capture the intriguing thought that "[i]f you know, you couldn't easily have been wrong" (Sainsbury 1997: p. 907). In other words, the safety-condition is meant to exclude cases in which a true belief is formed in a way that could have easily delivered error (cf. Hawthorne 2004: p. 56).

How is the discussion of modal conditions on knowledge connected to evolutionary debunking? Safety and sensitivity are often thought to be necessary conditions for knowledge, but EDAs attack epistemic justification. So how are the modal conditions relevant here?

The basic idea is that although safety and sensitivity are two conditions, which are often thought to be necessary for knowledge, it seems plausible to assume that receiving information that a belief of yours is insensitive or unsafe can defeat your justification for holding it. Indeed, the argument due to Justin Clarke-Doane, which will be presented in chapter (2), comes with an account of undercutting defeat that draws on both sensitivity and safety. As we will see there, Clarke-Doane assumes that if you have reason to think that your belief that p is either not safe or not sensitive, then this suffices for taking away your justification for believing that p.

Let me end this section by sketching out the rationale for this thought.[40] Let's first consider how receiving information that indicates insensitivity might take away

---

[39] In appendix B, I discuss these conditions and various formulations of them in more detail.
[40] This is a rather preliminary sketch, we will discuss the epistemology of Clarke-Doane's argument in much more detail in the chapters to come.

justification in the context of evolutionary debunking. Assume that our moral beliefs, produced by our moral faculty M,[41] are initially justified. Now, an evolutionary explanation might be thought to defeat our moral beliefs via showing those beliefs to be *insensitive*. That means this explanation now gives us reason to believe that there is a disconnect between our moral beliefs formed via M and the moral truths by giving us reason to think that the following claim is true:

> ***MORAL INSENSITIVITY***. If the moral truths were different, our moral beliefs would have been the same, i.e. in nearby possible worlds the moral truths are different, and our moral beliefs formed via M are thus false there.

MORAL INSENSITIVITY presupposes that the moral truths could have been easily different.[42] If evolutionary explanations give us reason to believe that MORAL INSENSITIVITY is true, then these explanations might defeat the justification of our moral beliefs via giving us reason to think that we could easily have ended up believing the same things we actually believe, even if the moral truths were different. And, intuitively, this seems to be an epistemically worrisome discovery.

And let's next consider how evolutionary explanations might defeat the justification of our moral beliefs via indicating that our beliefs are not modally safe from error. Evolutionary explanations might take away our justification by giving us reason to believe that the following claim is true:

> ***MORAL UNSAFETY***. Employing our moral faculty M could easily have led us to form different moral beliefs, while the moral truths remain the same. That means in nearby possible worlds employing M gives us divergent doxastic results, and these beliefs are false, since the moral truths stay fixed.

Here the information we receive suggests that the evolutionary influence on our moral belief-forming mechanism could have easily led us to hold different moral beliefs, even if the moral truths were to remain as they are in the actual case. Intuitively, our

---

[41] In this section, I bracket the question of what the relevant method or basis is in the case of our moral beliefs. See appendix C.

[42] Given what we have said in section (1.3.4), we can perhaps already tell how Clarke-Doane will try to defuse the defeating potential of MORAL INSENSITIVITY, but let's not get ahead of ourselves.

initial justification here seems to be defeated by the recognition that M does not provide the beliefs it produces with enough protection against error in similar cases.

Clarke-Doane's AMS now assumes that the *only way* for evolutionary considerations to undercut our moral beliefs, is via giving us reason to think that MORAL INSENSITIVITY and/or MORAL UNSAFETY are true. This assumption, which he calls "Modal Security", will be presented in chapter (2) and will be critically discussed in chapters (3) and (4).[43]

1.4 EVOLUTION AS AN EXPLANATORY CHALLENGE

In this section, I construct a generic EDA. As its name implies, the *generic* EDA is hardly original. In fact, this generic argument is an only slightly modified version of Dustin Locke's (2014) EDA.

The main purpose of this generic EDA in the context of this thesis will be twofold. First, I want to identify an argument that (epistemologically at least) qualifies as a serious competitor for further consideration in the philosophical debate. This argument deserves attention because it rests on a clear and plausible epistemological story. It is important to have such an argument in place, as EDAs (like the one's presented by Street and Joyce) have often been criticized for being "all murky where it counts most: exactly which epistemic principle combines with the facts of evolution to undermine moral realism?" (Bogardus 2016: p. 638; cf. also Berker 2014, White 2010; Brosnan 2011; Clarke-Doane 2012, Hanson 2017, Mogensen 2014). Here it is not my ambition to show you that the generic EDA is successful in refuting NNMR. My modest ambition is just to have an argument in place that is epistemologically clear and respectable.

Secondly, this argument is meant to provide a foil for the two influential responses to EDAs, third-factor accounts and the AMS, which are the main focus of this thesis. These two responses will be presented and critically discussed in chapters (2-4).

---

[43] For another modal account of undercutting defeat, similar in spirit to Clarke-Doane's, cf. Pollock (1987).

1.4.1 AN ACCOUNT OF UNDERCUTTING DEFEAT

The epistemology of evolutionary debunking is a controversial topic (cf. e.g., Bogardus 2016; White 2010). Thus, I suggest the following cautious way of proceeding. Let us first start with a popular and plausible general conception of undercutting mental state defeat. Then we can reflect on how to utilize this general notion of undercutting defeat in a generic EDA. So, before we ask what makes evolution a defeater, we should try to answer the question of what makes a defeater a defeater.

Before we start, I need to clarify a bit of terminology. I will repeatedly use the term "reliability" in this section. Obviously, given the importance of this term to contemporary epistemology, trying to get clear on what *reliability* is goes way beyond the scope of this thesis. Instead of getting us bogged down in difficult epistemological issues,[44] that I can't solve, I propose we use the following simple account to give us some guidance: I use the term "reliability" as meaning roughly (i) that the relevant belief-forming process/source/ method/mechanism/faculty of a subject S produces mostly true over false beliefs in the appropriate environment, and (ii) the fact that this belief-forming process/source/ method/mechanism/faculty produces mostly true over false beliefs in the appropriate environment is itself not a freaky coincidence. Conditions (i) and (ii) are admittedly vague, and I don't take them to be appropriate endpoints for any respectable account of reliability. But they are often taken as starting points (cf. Goldman 1979; Goldman & Beddor 2015: section 2), and I think that they are fine for the purposes of this thesis. This vague and intuitive notion can then be cashed out in various ways by epistemologists, e.g. in terms of modal safety from error.[45]

---

[44] Cf. appendix D for some cursory, but hopefully helpful remarks about the use of the term "reliability" in contemporary epistemology.

[45] Reliability is strictly speaking a property not of beliefs (or other doxastic attitudes) themselves, but of the relevant belief-forming process/method/mechanism/ basis/source /faculty of a subject S that produces those beliefs (or other doxastic attitudes). Sometimes I am going to write a bit loosely and say e.g., "that a belief is reliable" or that a "belief was formed reliably". But this should always be taken to mean that this belief was formed via applying a belief-forming process/method/mechanism/faculty that is reliable. I am also going to use the words "reliable" and "(epistemically) trustworthy" interchangeably when I refer to a belief-forming process/method/mechanism/ basis /source /faculty.

I will also make heavy use of the term "explanation". As with "reliability" in epistemology, there are a myriad of conceptions of "explanation" in the philosophy of science. Here, I follow Kevin McCain in taking a very general and inclusive approach towards explanation:

> The gist of this approach is that "explanations track dependence relations" of all kinds – causal relations, mereological relations, the relation of constitution, and so on. More simply, an explanation in this context can be understood as a set of propositions that provide an answer to a why-question, e.g. why did some event occur, why is some claim true, or why does something have the features that it does? (McCain 2016: p. 9)

Now to begin, let us take a step back and ask a very basic question about undercutting defeat:

> Why is it that receiving clear and convincing information that your way of arriving at a given belief that p is unreliable takes away the justification for believing that p?

A popular answer to this question builds on the idea that it is a requirement on your justifiably believing that p that you do not take or that you are not in a position to take a negative perspective toward whether your belief that p has been produced in an epistemically sound way (Moon 2018: pp. 256-257; cf. Bergmann 2005: p. 422).[46]

This requirement has only recently been made explicit by Andrew Moon (2018: p. 256), but I think it is safe to say that versions of this condition are widely endorsed.[47] Here I present a slightly amended version of Moon's formulation. Let "p*" denote the proposition that S's belief that p was formed reliably (via the method that S employed to form it) (cf. Bergmann 2005). Let S's "doxastic/normative set of background beliefs" denote the set of beliefs of S including both (i) the whole set of S's other beliefs and (ii) the set of beliefs that S should have on account of the evidence available to her (regardless of whether she actually has those beliefs). Defining S's relevant set of background beliefs in this way of course references the

---

[46] Borrowing a term from Alexander (2017: p. 900) the idea here seems to be that justification has a "perspectival dimension", i.e. the thought that a belief is justified only if it is permissible from one's perspective.

[47] For an impressive list of supporters cf. Moon (2018: p. 270).

notions of doxastic and normative defeat introduced in section (1.3.2) above. Then we have the following "no defeater"-requirement on S's justified belief that p:

> ***NO DEFEATER.*** S justifiably believes that p only if it is not the case that in virtue of S's doxastic/normative set of background beliefs, S should withhold or disbelieve p*.[48]

So, if S justifiably believes that p it is necessarily not the case that in virtue of her background beliefs or in virtue of beliefs that S should have on account of the evidence available to her, S should withhold[49] or disbelieve p*. NO DEFEATER basically tells us that for S's belief that p to be justified, it must not be the case that S has a doxastic or normative undercutting defeater for her belief that p. NO DEFEATER also gives us a partial account on how receiving clear and convincing information that your way of arriving at your belief that p is unreliable defeats your beliefs justification. If this information makes it the case that in virtue of your background beliefs or in virtue of beliefs that you should have (on account of the evidence available to you), you should withhold or disbelieve p*, then this generates an undercutting defeater for your belief's justification.

This condition is in line with the framework for epistemic defeat developed by Michael Bergmann (2005; 2006: Ch. 6). In Bergmann's view, if a subject has or should have a certain doxastic attitude (i.e., suspension of judgment or outright disbelief) towards the higher-order proposition p*, then this generates an undercutting defeater for her belief that p (Bergmann 2005: pp. 426–427). On this picture, these defeaters are always "mental state-defeaters": i.e., what defeats S's justification for believing that

---

[48] Cf. appendix E for some further clarifications of and motivation for NO DEFEATER.

[49] Throughout, I am going to use "withhold" as shorthand for "withhold belief in/ withhold from believing in". I also use "withhold" interchangeably with "suspension of judgment". As Locke (2014: p. 223) points out in a footnote, obviously withholding belief that p is not equivalent to believing that not-p, but neither is it equivalent to not believing that p, although this might be less obvious. Withholding belief is more active and implies that S has considered whether to believe that p. S might not believe that p, because she simply has never even considered whether p is true, but for S to withhold belief that p, it is necessary that S has considered this.

p is a doxastic attitude (i.e., suspension of belief or disbelief) that S has or should have. Within this framework, it is these mental states, which constitute defeaters.[50, 51]

Here is some intuitive motivation for NO DEFEATER (this is a case also taken from Moon 2018: p. 256). Let "CAR" denote the proposition that this is a red car and let "CAR*" denote the proposition that my belief that this is a red car was reliably formed. Johnny comes to believe CAR. In a first case, Johnny then comes to believe (or Johnny should come to believe) that before he arrived at his belief that CAR he has been administered medication that makes 99% of people hallucinate red cars. In a second case, Johnny then comes to believe (or Johnny should come to believe) that before he arrived at his belief that CAR he has been administered medication that makes 50% of people hallucinate red cars. In both cases, we can stipulate that Johnny does not actually hallucinate, he looks at a real red car in the driveway, and reliably forms the belief that CAR. Nonetheless, given Johnny's doxastic/normative set of background beliefs, it seems that Johnny should disbelieve CAR* in the first case, and that he should withhold CAR* in the second case.[52] Furthermore, in both cases, it seems that his belief in CAR is not justified. It seems that his doxastic/normative set of background beliefs defeats his justification for believing CAR via making it the case that he should withhold or disbelieve CAR*. Moreover, as Moon notes, this seems to be true in general:

> For any belief that p, if one should disbelieve p* or withhold p*, then it seems that one is not justified in believing p. (Moon 2018: p. 257)

---

[50] In what follows, I am sometimes going to write a bit loosely: e.g., I might write that "evolutionary evidence generates a defeater" or that "a piece of information constitutes a defeater". This should always be understood as meaning that the relevant piece of evidence or information generates a defeater by (and only by) making it the case that S should now withhold or disbelieve that p*. Strictly speaking, the defeater is not the evidence or information that S receives, but the doxastic attitude towards the higher-order proposition p* that S forms or should form in response to that evidence or information.

[51] Does the focus on "mental state" defeaters bias the above account of defeat towards epistemic internalism? No: see appendix A. Cf. also appendix E.

[52] Here I assume that what doxastic attitude a believer should take towards a proposition at least partly depends on how probable a believer thinks (or should think) it is that this proposition obtains.

As I have noted, variations of NO DEFEATER seem to be rather widely endorsed. NO DEFEATER can be amended in various ways, to suit your epistemological convictions.[53]

1.4.2 AN EXPLANATORY DEMAND

The explanatory EDA I develop here (following Locke 2014: pp. 220-221, 223-225, 227-232) now seizes on an interpretation of what makes it the case that S *should* withhold or disbelieve p* by formulating an explanatory demand on what we can continue to believe with justification. Before going into a bit more detail on this interpretation, it might be useful to dwell briefly on the intuitive basis for such a demand, and to sketch out how this demand figures in Locke's EDA.

This explanatory demand is based on the intuitive thought that it is often the case that I will not be able to continue to justifiably believe that p if the truth of my belief that p does not figure in the best explanation available to me for why I believe that p. For example, we might, intuitively, doubt that my belief that p is still justified, if I am now presented with clear and convincing information that makes it the case that the best explanation available to me for why I have come to believe that p nowhere presupposes, posits, entails or makes likely the truth of my belief.

When we apply this thought to our own belief-forming mechanisms, we get a demand that tells us that if the best explanation (available to us) of our belief-forming mechanism in no way involves the truth about p, then we should withhold (or disbelieve) whether this mechanism reliably produces beliefs about p.

Now, Locke's EDA features an epistemic principle of just this kind. This principle tells us that for it to be the case that we should withhold p* it suffices if we should withhold belief in the claim that the explanatory history of the relevant belief-forming mechanism does not involve the truth of the beliefs it produces. Locke then uses this principle to argue that if we have very good reason to believe an evolutionary explanation of our moral belief-forming faculty that nowhere involves non-naturalistically construed moral truths or facts, then this means that we should

---

[53] Moon (2018: pp. 257-258) sketches out a few of these ways. For example, how one applies NO DEFEATER, will partly depend on one's account of S's available "evidence" (e.g., S's beliefs or S's true beliefs or S's justified beliefs or S's justified true beliefs or S's knowledge…).

withhold belief about the claim that the explanatory history of our moral belief-forming faculty (and thus also of our moral beliefs) involves moral truths or facts. Since withholding belief about the claim that the explanatory history of our moral belief-forming mechanisms involves moral facts undermines our initial justification for believing that our respective faculty is reliable, evolution potentially provides us with an undercutting defeater for our moral beliefs.

That's the rough picture of Locke's EDA. Now I want to go into a bit more detail. Let's discuss what makes it the case that S should withhold or disbelieve p* on this interpretation. Here's the principle that Locke (2014: p. 232) presents and that together with another condition that I will discuss in a moment gives us a sufficient condition on when S should withhold p*. For a subject S, let M be one of S's cognitive mechanisms to form beliefs involving the concept X, and let "X-facts" refer to facts involving the object/property/relation/kind/etc. picked out by X:

> ***COGNITION DEFEAT***. If S withholds or should withhold belief that the explanatory history of M involves X-facts, then S's initial justification to believe that M is reliable is lost.

Suppose that S was initially justified to believe that her belief that p[54] has been reliably produced by M, i.e. S was justified to believe p*. Initially, it was not the case that given her doxastic/normative set of background beliefs she should have withheld or disbelieved p*. Perhaps, S's initial justification is in part due to a default entitlement to belief in the reliability of one's own cognitive mechanisms. That our cognitive mechanisms are innocent until proven guilty is a highly useful assumption for combatting global sceptical arguments. As Locke points out, in contrast to globally sceptical challenges the above formulation of COGNITION DEFEAT is entirely compatible with this assumption (2014: pp. 230-232). S's initial justification for believing p* is only lost after S receives information that is apt to have this effect.

Say that S receives information that makes it the case that she should withhold belief that the explanatory history of M involves X-facts. COGNITION DEFEAT states

---

[54] Where S's belief that p formed via M involves the concept X. Here is an example: Say "M" is S's moral faculty. This faculty produces beliefs involving the concept of duty. Joanna's belief that Max has the duty to care for his aging grandmother is formed via Joanna's moral faculty.

that this suffices for S to lose her initial justification for believing p*. Assume for the moment that S also lacks any other source of justification for her belief that p* that is independent of M. In that case, it seems that S's belief that p* given her updated doxastic/normative set of background beliefs is unjustified. In section (1.3.2), we have observed that if you are not justified in believing in a claim, you are not permitted to believe that claim, and if you are not permitted to believe that claim, then you ought not to believe it. Therefore, if you become unjustified in believing that p*, then at very least, you should withhold p*. Given NO DEFEATER, this generates a defeater for the justification of S's belief that p: You are not justified to believe that p if you should withhold or disbelieve that p*.[55]

There's still one question left to settle in this section: When should S withhold belief that the explanatory history of M involves X-facts? Here's the answer I favour:

> **BEST EXPLANATION**. If the best explanation available to S of her belief-forming process M nowhere involves (=presupposes, posits, implies or makes likely) the existence of X-facts, then S should withhold belief about the claim that the explanatory history of M involves X-facts.

BEST EXPLANATION strikes me as the intuitively most plausible answer to the above question.[56]

1.4.3 INDEPENDENT MEANS OF CONFIRMING RELIABILITY

COGNITION DEFEAT states a sufficient condition on when S loses her initial justification for believing that p*. However, losing your initial justification for believing a certain claim is not equivalent to being unjustified in believing that claim. Let us assume that it is indeed the case that S loses her initial justification for believing p* by withholding that M's explanatory history involves X-facts. But perhaps S can gain new justification for believing p* by finding the means to independently confirm that M reliably produces beliefs that p (Locke 2014: p. 232). This would enable S to reinstate her justification for believing that p*. Following Locke, we can thus

---

[55] For further motivation for COGNITION DEFEAT, see appendix F.

[56] For further discussion of and motivation for BEST EXPLANATION, please see appendix G.

formulate the following necessary condition for when S can still be justified in believing p*, while withholding that M's explanatory history involves X-facts:

> ***INDEPENDENCE CONSTRAINT***. If S withholds belief that her belief-forming mechanism M has an explanatory history that involves X-facts, it is permissible for her to believe that M is reliable only if S's belief that M is reliable is based on a source for justification that is itself independent of M.

COGNITION DEFEAT tells us that if S withholds belief that the explanatory history of M involves X-facts, then S's initial justification to believe that M is reliable is lost. The INDEPENDENCE CONSTRAINT now spells out a necessary condition on when you can withhold belief in the claim that the explanatory history of M involves X-facts without losing your justification: only if you have M-independent means of showing that M is trustworthy.[57, 58]

1.4.4 THE EPISTEMOLOGICAL UNDERPINNING

We are now able to tell a reasonably comprehensive and seemingly plausible story concerning the epistemological underpinning of the explanatory challenge to non-naturalist moral realism under consideration.

Suppose that S was initially justified to believe that her belief that p has been reliably produced by M, i.e. S was justified to believe p*. Initially, it was not the case that given her doxastic/normative set of background beliefs she should have withheld or disbelieved p*. S then receives information that makes it the case that the best explanation available to S about her belief-forming process M nowhere presupposes, posits, implies or makes likely the existence of X-facts. If the best explanation available to S concerning M nowhere presupposes, posits, implies or makes likely the existence of X-facts then then S should withhold belief about the claim that the explanatory history of M involves X-facts.

If S should withhold belief about the claim that the explanatory history of M involves X-facts, then this suffices for S to lose her initial justification for believing p*. If S withholds belief that her belief-forming mechanism M has an explanatory history that involves X-facts, it is permissible for her to believe that M is reliable only if S's belief

---

[57] Lutz's (2017) EDA features a similar demand for independence.

[58] For motivation for the INDEPENDENCE CONSTRAINT, please see appendix H.

that M is reliable is based on a source for justification that is itself independent of M. Assume that S's belief that M is reliable cannot be based on a source of justification that is itself independent of M.

If it is not epistemically permissible to believe that M is reliable, then S's belief that M is reliable is unjustified. In other words, S's belief that p* is unjustified given her updated doxastic/normative set of background beliefs. If you are not justified in believing in a claim, you are not permitted to believe that claim, and if you are not permitted to believe that claim, then you ought not to believe it. So, if S is not justified in believing that p*, then S is not permitted to believe and ought not to believe p*. Therefore, if S becomes unjustified in believing p*, then at the very least, S should withhold p*.

S's belief that p is justified only if it is not the case that in virtue of her doxastic/normative set of background beliefs S should withhold or disbelieve p*. Therefore, if S should withhold p*, this generates an undercutting defeater for the justification of S's belief that p. This is the whole epistemological story underpinning the generic EDA.[59]

1.4.5 A GENERIC EXPLANATORY EDA

Finally, we are in a position, to formulate a generic EDA that seems epistemologically clear and respectable, and that can serve as a foil for the standard responses discussed in the following chapters. Now for the argument to get off the ground, we need to make two empirical assumptions (Locke 2014: p. 220 & p. 223):

(I) Our moral faculty[60] is (largely) the product of our evolutionary history.

(II) Our moral faculty was selected because it had some natural property N (e.g., it contributed to reproductive success by promoting certain kinds of cooperation amongst our ancestors).[61]

---

[59] For another succinct statement of this story, and for some further motivation, please see appendix I.

[60] Where the "moral faculty" is a psychological mechanism that is crucially involved in the production of many or most of our moral beliefs. This mechanism could e.g., produce moral intuitions, which regularly give rise to corresponding moral beliefs. For more on the moral faculty, see appendix C.

[61] Needless to say, these assumptions are controversial, cf. Kahane 2011; Fitzpatrick 2015; Isserow 2018. Cf. also the discussion of the idealizing assumptions in appendix C.

Under the assumption that there is strong evidence for (I) and (II),[62] it seems we now have what we need to get a generic explanatory EDA off the ground. The debunker now has what looks like a good case for the claim that the best explanation available to a moral believer who is confronted with the evidence for (I) and (II) nowhere involves the existence of non-naturalistically construed moral facts. Applying the above epistemological story, we now can construct an argument that generates a defeater for our moral beliefs (provided we assume NNMR):

Generic EDA:

(1) The best explanation available to you for your moral faculty nowhere involves (=presupposes, posits, implies or makes likely) the existence of moral facts.

(2) If the best explanation available to you concerning your moral faculty nowhere involves the existence of moral facts, then you should withhold belief about the claim that the explanatory history of your moral faculty involves moral facts.

(3) Therefore, you should withhold belief about the claim that the explanatory history of your moral faculty involves moral facts.

(4) If you should withhold belief about the claim that the explanatory history of your moral faculty involves moral facts, then it is epistemically permissible for you to believe that your moral faculty is reliable only if this belief is based on a source for justification independent of your moral faculty.

(5) It is epistemically permissible for you to believe that your moral faculty is reliable only if this belief is based on a source for justification independent of your moral faculty

(6) Your belief that your moral faculty is reliable is not based on a source for justification independent of your moral faculty.

(7) It is not epistemically permissible for you to believe that your moral faculty is reliable.

(8) If it is not epistemically permissible for you to believe that your moral faculty is reliable, then you should withhold or disbelieve that your moral faculty is reliable.

---

[62] And provided that the idealizations made in appendix C hold.

(9) You should withhold or disbelieve that your moral faculty is reliable.

(10) If you are justified in holding your moral beliefs formed via your moral faculty, then it is not the case that you should withhold or disbelieve that your moral faculty is reliable.

(11) You are not justified in holding your moral beliefs formed via your moral faculty.[63]

(1-10) comprise the application of the abstract epistemological story above to the evolutionary explanation of our moral faculty. Assume that you have sufficient epistemic reason to believe (I) and (II). The truth of (I) and (II) (plus the idealizing assumptions (a-c) presented in appendix C) furnishes you with an evolutionary explanation of your moral faculty that nowhere involves the existence of moral facts. Assume that this explanation concerning your moral faculty is better than any rival explanation of your moral faculty. In other words, it is better than any explanation of your moral faculty that involves the proposition that moral facts exist. This supports (1). (2) is just the application of BEST EXPLANATION. (3) follows from (1) and (2). (4) states the application of the INDEPENDENCE CONSTRAINT. (5) follows from (3) and (4). (6) states that your belief that your moral faculty is reliable does not satisfy the INDEPENDENCE CONSTRAINT. The reason is simple: (given the idealizations of appendix C) all your moral beliefs are the product of your moral faculty. (7) follows from (5) and (6). (8) seems clear as day, and has been argued for in section (1.4.4) above. (9) follows from (7) and (8). (10) is based on NO DEFEATER: if it is necessary for being justified in believing that p via M that it is not the case that you should withhold or disbelieve that M is reliable, then if you are justified in believing that p via M, it is not the case that you should withhold or disbelieve that M is reliable. (11) follows from (9) and (10). And (11) is just the kind of conclusion you would want to have as an evolutionary debunker.

---

[63] As NNMR is the view targeted by this argument, it is assumed that the "you" that this argument addresses shares the commitments of NNMR that give rise to this argument (although it is not necessary that the subject has a philosophically sophisticated grasp of these commitments). But that's fine: defenders of NNMR often claim that of all available meta-ethical positions, NNMR is the view that undergirds our common-sensical moral practices and discourses and the phenomenology of our moral experience (cf. Cuneo 2011; Enoch 2011; Kramer 2009). And since the argument is targeting NNMR, it is also fine if it does not work if the "you" happens to be e.g. a moral naturalist.

We now have a generic explanatory EDA based on a sufficiently stable epistemological foundation. This wraps up our task for this chapter. With this argument now in hand, we can go on to discuss and evaluate third-factor accounts and the AMS in the remaining chapters.

2 TWO STANDARD RESPONSES TO EDAS

2.1 INTRODUCTION

In the last chapter, we constructed an epistemologically respectable EDA. This chapter and the next two will be concerned with coming to grips with two standard responses to EDAs. In this chapter, I will present these responses, while in the following two chapters I will argue against them. Here is a preview of the argument that will emerge from chapters (2-4): Clarke-Doane and the third-factor theorists are ultimately confronted with the unattractive choice between conceding epistemic defeat or recommending an option that sanctions epistemically irrational doxastic behaviour. In any case, the AMS and third-factor accounts are thus shown to be bad responses to the generic EDA, as they fail to show that evolutionary considerations do not render us epistemically criticisable for holding or continuing to hold our moral beliefs.[64]

I use the term "standard responses" because (i) these two responses have been much discussed in the literature, and (ii) these responses are importantly similar in several respects. These responses are the argument from modal security (Clarke-Doane 2015 & 2016) and third-factor accounts (e.g., Enoch 2010 & 2011; Wielenberg 2010). Importantly, these responses do not focus on the empirical assumptions that the debunkers make, but rather argue that even if the debunkers' empirical story is largely correct, and even if we assume non-naturalist moral realism, the epistemic threat supposedly arising from evolution can be neutralized.[65]

---

[64] While we assume NNMR.

[65] In the philosophy of religion, Alvin Plantinga (2000) employs a strategy for dealing with what he calls "*de jure* objections" to theistic belief (=objections, which try to establish the point that theistic belief, whether true or not, lacks a positive epistemic property) which is eerily similar to the standard responses. *De jure* objections include debunking explanations of theistic belief. In response to *de jure* objectors, Plantinga argues for the following conditional conclusion (2000: 189-190):

> *Plantinga's Conditional Conclusion*. If God exists, then belief in God is probably *prima facie* justified.

A powerful implication that Plantinga draws from this conditional conclusion is that since theistic belief probably has justification if it is true, if you want to raise objections to the justification of theistic belief, you must show that the antecedent in *Plantinga's Conditional Conclusion* is false.

These arguments share several important similarities. Both the AMS and third-factor accounts suppose that the epistemic threat in need of neutralizing is roughly that evolutionary considerations might show our moral beliefs to be subject to a problematic kind of epistemic luck. Both Clarke-Doane, as well as the third-factor theorists think that the potential epistemic danger lies in that an (i) evolutionary explanation of our moral belief-forming mechanisms, in conjunction with (ii) a commitment to a picture of moral facts as causally inert and causally and constitutively independent of our moral attitudes seems to imply (iii) that it would be a "cosmic coincidence" if our moral beliefs were actually true (cf. Bedke 2009). Both Clarke-Doane and the third-factor theorists think that this sort of coincidence could potentially undercut the justification of our moral beliefs.

Another important similarity is that in trying to neutralize this potential epistemic danger via arguing for the conclusion that it is not a coincidence that our moral beliefs land on the truth, both responses rely on roughly the following claim:

> **COMMON-SENSE MORALITY**. Our common-sense moral beliefs are by-and-large true.[66]

And the third important similarity is that both the AMS and third-factor accounts utilize COMMON-SENSE MORALITY in the same fashion: they argue that if our moral beliefs are true, then they are reliably formed – and evolutionary considerations cannot give us a reason to believe differently. In short: they argue that if our beliefs are true, and provided some other conditions obtain, then evolutionary consideration do not undercut the justification of our moral beliefs.

If true, this means that you cannot call into question the epistemic justification of theistic belief without calling into question its truth. This renders the whole family of *de jure* objection impotent without the support of a successful *de facto* objection.
I tend to think that versions of the arguments developed in the next few chapters should also apply to Plantinga's strategy for dealing with *de jure* objections, but supporting this point is a project for another occasion.

[66] As I have mentioned in the introduction, third-factor responses and the AMS belong to a family of responses to EDAs that have recently been called "first-order replies" by Morton (2018: p. 2), and "minimalist replies" by Locke (2014: p. 221, p. 227). These are replies which all crucially assume substantive moral claims. I lack the space for showing this, but I tend to think that a version of the arguments developed in the following chapters should also apply to the members of this family, which I do not discuss in this thesis.

Both the AMS and third-factor accounts are therefore crucially similar in certain key respects. In the next chapter, we will see that relying on the truth of our moral beliefs in the face of the evidence or information involved in the generic EDA from the last chapter is problematic.

Here is how I will proceed in this chapter. In (2.2) I will introduce the argument from modal security. In section (2.3), I will present third-factor accounts. In (2.4), I will summarize the most important points.

2.2 THE ARGUMENT FROM MODAL SECURITY

In this section, I want to give you a succinct presentation of the AMS due to Justin Clarke-Doane (2012; 2015; 2016). In a nutshell, this argument aims to show that if some of our basic moral beliefs are true, and if we could not have easily failed to have these beliefs, then the epistemic threat supposedly arising from evolutionary considerations can be neutralized.

Like third-factor accounts, to provide a defence of our common-sense moral beliefs against the debunking threat, the AMS needs to assume that our basic moral beliefs are mostly true. The assumption that our common-sense moral beliefs are mostly true is crucial, since the AMS aims to establish the conditional claim that if our basic moral beliefs are true, then they are reliably true. And so, to get to the conclusion that our moral beliefs are reliable, the AMS needs to assume that the antecedent of that conditional is true.

Is it legitimate for the AMS to assume our basic moral beliefs are true? This will be a main topic of the following chapters, so I will postpone discussing this question until then. Clarke-Doane points out that since EDAs do not provide us with rebutting defeaters, they do not give us any direct reason to think that any moral belief of ours is wrong. He therefore holds that it is compatible with EDAs that there are moral truths, some of which might be believed by us. We will return to this point below.

It is worth pointing out that the scope of the argument is restricted to only directly protecting our justification for believing basic moral truths from arguments like the generic EDA. However, Clarke-Doane also argues that once we have secured justification for belief in the basic moral truths, then it might be reasonably expected

that we are able to justifiably infer non-basic moral truths from them (Clarke-Doane 2016: p. 29, p. 35).

So, in the first instance, the AMS needs to assume that our basic moral beliefs are true. Furthermore, Clarke-Doane relies on the claim that the basic moral truths (if they exist) are metaphysically necessarily true. The motivation for this assumption has been presented in section (1.3.4). Recall: The metaphysical necessity of the basic moral truth is part of a strong supervenience-claim. The necessity-claim is then well motivated if the supervenience-claim is. Motivation for the strong supervenience-claim is provided by the fact it gives us a general account about the relation between two families of properties (moral and non-moral) that rules out the possibility mentioned above that there could be two cases, which are identical with respect to their non-moral properties, but which differ with respect to their moral properties. (This motivation of course assumes that some moral properties are instantiated, this is another reason why Clarke-Doane needs to rely on COMMON-SENSE MORALITY.) Let's assume that the metaphysical necessity-claim is indeed true.

Now, we can look at the substantial epistemological story on which the argument rests. Clarke-Doane assumes the following account of undercutting defeat:

> ***MODAL SECURITY***. Information, E, cannot undermine our beliefs of a kind, D, without giving us some reason to believe that our D-beliefs are not both safe and sensitive.[67]

That means for E to take away your justification for believing that p, E needs to give you a reason to think that your belief that p is either not safe or not sensitive. MODAL SECURITY will be called into question in due course, but in this chapter, I focus

---

[67] Clarke-Doane (2015: p. 97) formulated this condition to cover both undercutting and rebutting defeaters. The reason supporting this broader formulation is that a rebutting defeater also gives you a reason that your relevant belief is not both safe and sensitive by indicating that the belief is false. This is due to the fact that a false belief can neither be safe nor sensitive (ibid.). Clarke-Doane (2016: p. 31) recants this. The reason for this seems to be that while every defeater (in Clarke-Doane's book) indicates that a belief is either not safe or not sensitive, undercutting defeaters defeat a belief's justification *only* via indicating that it is either not safe or not sensitive (as opposed to defeating it via indicating its falsity *and* its lack of safety or sensitivity).

exclusively on presenting the standard responses. Critical discussion will take place in the next two chapters.

And to give us a reminder, here are the final versions of the principles of safety and sensitivity from section (1.3.5):

> ***Sensitivity$_3$***. S's belief that p formed via or based on M is sensitive if and only if: in nearby possible worlds where p is false, S would not believe that p via or based on M.
>
> ***Safety$_3$***. S's belief that p formed via or based on M is safe if and only if: in nearby possible worlds where S believes that p via or based on M, p is true.

So, if evolutionary considerations do not undermine the justification for our basic moral beliefs, it must be the case that these considerations do not give us a reason to think that our basic moral beliefs are either not safe or not sensitive. Why is that so?

Let's start with looking into the safety of our basic moral beliefs first. Here we encounter a complication. Given that the fundamental moral truths are necessary and given that our belief has as its content a basic moral truth, then there's no close possible world where we have this belief while it is false, since there is no possible world where the relevant belief is false. Does this mean that belief in necessary truths is safe by default?

No. Epistemologists who hold that safety is a necessary condition for knowledge have argued that we can nonetheless evaluate beliefs that have a necessary truth as their content in terms of safety. These epistemologists hold that the safety-condition should be formulated in such a way as to ensure that a belief formed via the relevant method is "safe from error". For your belief to be safe, it does not suffice that *this belief* could not easily have been wrong. It must also be true, that M could not easily have given you a false belief in a different proposition. For example, Duncan Pritchard writes that "[all] we need to do is to talk of the doxastic result of the target belief-forming process, whatever that might be, and not focus solely on the belief in the target proposition" (Pritchard 2009: p. 34). And in a similar vein, Timothy Williamson writes that an agent's belief in a necessary truth can still be unsafe and thereby fail "to be knowledge because the method by which he reached it could just as easily have led to a false belief in a different proposition" (Williamson 2000: p. 182).

Thus our basic moral beliefs could still be unsafe, at least in the sense that although the contents of our present beliefs might be necessarily true, we easily might have believed differently while still using the same method, and so we could easily have had false beliefs.

So, to have a condition that can give us sensible results in cases where somebody forms beliefs in necessary truths in a bad fashion we must amend the safety-condition in a way that reflects Pritchard's and Williamson's comments. Max Baker-Hytch presents us with such an amended safety-condition:

> ***AMENDED SAFETY***. "S safely believes that p in world w via method M iff there is no nearby world w* in which S arrives at a false belief with a relevantly similar propositional content and with a relevantly similar causal history." (2014: p. 175)

AMENDED SAFETY can deal with epistemically fishy belief in necessary truths: Let's say that my belief that God exists is a belief in a necessary truth. Let's say I have formed this belief via the testimony provided by a person who unbeknownst to me is a habitual liar on matters of religious concern. This lying person decided to tell me that the God of "Perfect Being-Theology"[68] existed on a whim, but he also might have easily told me that Zeus exists. If he had told me that Zeus exists, I would have believed it. Although this time around I may have gotten the correct answer by consulting this person, there will be some nearby worlds in which I form a distinct but similar false belief, a false belief with a very similar causal history to my actual belief. My belief that God exists is therefore unsafe, even though it is a belief that has a metaphysically necessary proposition as its content.

Given that our basic moral beliefs have metaphysically necessarily truths as their contents, there are no relevantly nearby possible worlds where these same basic moral beliefs are false, since there is no possible world where they are false. So, with the AMENDED SAFETY-condition in hand, we need to ask whether there are nearby worlds where we have different moral beliefs (Faraci Msc: p. 4). Holding fixed the actual types of mechanisms bringing about our moral beliefs and the actual moral truths, do evolutionary considerations give us reason to think then that we could easily

---

[68] Cf. section (3.3.1), FN 82.

have come to have different and false moral beliefs? This seems to be affirmed by e.g. Braddock, Mogensen and Sinnott-Armstrong when they write:

> [D]ifferent instantiations of the process of cultural group selection have produced divergent normative systems, which nonetheless solve the same design-problem: namely, that of getting human societies to function as adaptive corporate units. In this way, one and the same process type may, through its various instantiations, easily result in divergent moral systems. (Braddock, Mogensen & Sinnott-Armstrong 2012)

But as Clarke-Doane argues, it is ironically the robustness of evolutionary explanations that could help to ensure the safety of our beliefs: Suppose again that we have the basic moral beliefs that we actually have for reasons that do not turn on their truth, but because, given our evolutionary history, these beliefs raised our reproductive fitness. Evolutionary forces would have made it the case that we would have these beliefs, no matter the truth. But if evolutionary forces would have made it so that we have these beliefs no matter the truth, then it seems these beliefs could not have been easily different. And so,

> given that our moral beliefs are actually true…and that the…basic moral truths could not have been different, we could not have easily had false…basic moral beliefs (Clarke-Doane 2016: p. 29).

This argument does not mean to conclusively show that our moral beliefs are safe. Rather, Clarke-Doane's point here is that an argument for their safety can be made that does not rest on our ability to explain their correspondence with the truth in any other sense (Faraci Msc.: p. 4). And so, this explanation seems compatible with, e.g. explanations of our moral belief that do not involve the relevant moral truths.

It should be no surprise that it is controversial how modally robust evolutionary explanations really are. Baker-Hytch, who in an unpublished manuscript develops a line of argument similar to the one supported by Clarke-Doane, points to Street for support for the claim that "evolution pushes human-like creatures in the direction of a certain fairly narrow range of biologically optimal moral belief-forming

[tendencies]" (Msc.: pp. 17-18).[69] This leads Baker-Hytch to conclude, that for at least many basic moral beliefs it is true that

> the method or causal process as a result of which we acquired our moral beliefs is such that it results in pretty much the same beliefs in nearby possible worlds… [therefore,] if… [our actual] beliefs are true, then the beliefs that that method produces in nearby possible worlds are also true. (ibid.)

Joyce (2016b: pp. 131-132) on the other hand disputes that evolutionary explanations are robust enough to establish the strong claim that our evolutionary history makes it the case that we could not have had different basic moral beliefs. He supports this by pointing to the actual diversity of moral beliefs across human cultures. This diversity will make it difficult to find a plausible candidate for any moral belief that is biologically entrenched in most individuals across most cultures. Furthermore, he argues that only an extreme (and in turn extremely implausible) version of moral nativism[70] would hold or support the claim that *particular* moral beliefs would have been selected for.

Nevertheless, supposing that this implausible version of moral nativism is true, seems for Joyce the best shot for supposing that evolution does supply us with an appropriately robust explanation for why we hold our basic moral beliefs. If the AMS must rely on an extreme version of moral nativism to establish the claim that our moral beliefs are safe, this would provide the debunker with an easy reply: argue that since this extreme moral nativism is likely to be false, we have good reason to think that evolutionary considerations can show our moral beliefs to be unsafe.

---

[69] Here is the relevant quote by Street:
"From an evolutionary point of view, these and many other of our basic evaluative tendencies are no accident. It is fairly obvious why, other things being equal, ancestors with these evaluative tendencies would have left more descendants than counterparts who, for example, viewed their survival as bad, their children's lives as worthless, or the fact that someone has helped them as a reason to hurt that person in return." (Street 2008: p. 208)

[70] We have come across moral nativism in section (1.2.2). Here is a reminder of what it is: moral nativism holds that humans have a specialized innate mechanism (or mechanisms) that accounts for their tendency to categorize the world in moral terms.

Perhaps, Clarke-Doane could look to the philosophy of religion to find another way of defending the safety of moral belief. More specifically, he could perhaps look to Baker-Hytch's defence of the safety of theistic belief (2014: pp. 178-180).

Baker-Hytch's reasoning there has the following starting point: a notorious problem with modal conditions in epistemology is that crucial notions like the "closeness" of possible worlds and "relevant similarity" are so vague. One way to deal with this problem is to use our intuitive judgments about cases on which everyone agrees to find an "anchor"-case. What's meant by "anchor" here? We start with a case where we all agree that it constitutes knowledge, e.g. an ordinary case of testimony-based belief. Then we start looking for the next closest error world (=possible worlds in which a subject S who correctly believes p via M in the actual world, believes a false proposition similar to p via M).

Following Baker-Hytch, one could now go on to argue that for some ordinary testimony cases, the closest error world is intuitively at least as close to the actual world, as it is for fairly typical cases of basic moral belief. One could then further argue that for the ordinary testimony-case, the causal history in the closest error world to the belief's causal history in the actual world is at least as similar as it is for typical cases of basic moral beliefs. As Baker-Hytch does for the case of typical instances of theistic belief, one could then contend that these two points are at least true for the person whose moral beliefs are assumed to be true, and whose moral belief is, as we could call it, "causally insulated" by e.g. her culture. That means, that for this person a considerable counterfactual alteration of her environment would have to take place for her to e.g. have been placed under the influence of a diverging moral system.

This results in a potential dilemma for the debunker who wants to draw on the contingency of moral belief to show that our moral beliefs are unsafe: either (i) the argument is only successful if focused on moral beliefs of people whose circumstances are such that they easily could have had different beliefs. Or (ii) the safety-measure applied to undermine causally insulated moral belief will also undermine ordinary and perfectly fine cases of testimony-based belief.

This is obviously only a sketch of a possible argument for the safety of (some) moral belief of perhaps at least some people. Here, I am going to grant Clarke-Doane that

some argument of this sort is available to him to support his point that our basic moral beliefs are safe, if true. I think the AMS can be shown to be an unsatisfactory response to the generic EDA even if we grant this. So much for the argument for the safety of our true basic moral beliefs. Now let's turn to sensitivity.

Attempting to make use of the sensitivity-condition for debunking morality runs into a well-known problem that sensitivity has with beliefs who have as their content a necessary truth (cf. e.g., Faraci Msc.; Roland & Cogburn 2011). If a belief is metaphysically necessarily true, then the relevant counterfactual *if the truth had been different then our beliefs would have been correspondingly different* is trivially true across metaphysically possible worlds (at least on the standard semantics for counterfactuals). Therefore, our relevant belief is also sensitive by default because since there is no close possible world where the truth is different, but I still have the same belief, because there is no possible world where the truth would have been different.

If Clarke-Doane is right, then this means that at least some of our moral beliefs are safe and sensitive, the debunking arguments notwithstanding. Now, MODAL SECURITY basically tells us that there is no such thing as a non-modal undercutting defeater:

> ***MODAL SECURITY***. Information, E, cannot undermine our beliefs of a kind, D, without giving us some reason to believe that our D-beliefs are not both safe and sensitive.

In defence of MODAL SECURITY, Clarke-Doane writes that it is hard to see how information could undermine our beliefs without challenging their safety or sensitivity: after all, if (i) there is no rebutting defeater for our relevant beliefs, and (ii) our beliefs are modally secure, i.e. they are both safe and sensitive, then it seems they "were (all but) bound to be true" (Clarke-Doane 2016: p. 33).

And now let us put all the material in this section together, and state the argument from modal security:

THE ARGUMENT FROM MODAL SECURITY.

(i). Our basic moral beliefs are true.

(ii). If true, our basic moral beliefs are metaphysically necessarily true.

(iii). Therefore, our basic moral beliefs are metaphysically necessarily true.

(iv). If our basic moral beliefs are metaphysically necessarily true, then our basic moral beliefs are sensitive by default since there is no close possible world in which they are false.

(v). Therefore, our basic moral beliefs are sensitive.

(vi). If our basic moral beliefs are metaphysically necessarily true and if we could not easily have failed to have these beliefs, then our basic moral beliefs are safe.

(vii). We could not easily have failed to have our basic moral beliefs.

(viii). Therefore, our basic moral beliefs are safe.

(ix). Our basic moral beliefs are both safe and sensitive.

(x). Information, E, cannot undermine our beliefs of a kind, D, without giving us some reason to believe that our D-beliefs are not both safe and sensitive.

(xi). If our basic moral beliefs are true, then evolutionary explanations of our moral beliefs do not give us a reason to believe that our basic moral beliefs are not both safe and sensitive.

(xii). Therefore, evolutionary explanations do not provide us with undercutting defeaters for our moral beliefs.

(xiii). Evolutionary explanations do not provide us with rebutting defeaters for our moral beliefs.

(xiv). Therefore, evolutionary explanations neither undermine, nor rebut our basic moral beliefs.

(xv). The justification of our basic moral beliefs is not defeated by evolutionary explanations.

(xvi). We are able to justifiably infer non-basic moral truths from our basic moral beliefs and our non-moral background knowledge.

(xvii). So, evolutionary explanations do not defeat the justification for all of our basic and non-basic moral beliefs.

2.3 THIRD-FACTOR ACCOUNTS

Another kind of standard response to EDAs in the above sense are third-factor accounts. Before we can discuss these responses themselves though, we need to provide some context.

In section (2.2), we have seen that Clarke-Doane thinks that the debunkers aim (or in any case, would need to aim) at proving that evolutionary considerations show our beliefs to be either not safe or not sensitive. In a similar vein, third-factor theorists, seem to be worried that evolutionary considerations could give us reason to think that even if our moral beliefs were true, we would have arrived at the truth only due to a problematic sort of epistemic luck.

How might an evolutionary explanation of our moral beliefs imply that these beliefs if true, are true only as a matter of epistemic luck? The idea here seems to be that it might require an extraordinary coincidence if evolution has favoured the production of true moral beliefs given that matters of moral truth and falsity are irrelevant in accounting for the selection pressures that shaped human moral psychology. This idea is clearly recognizable in Street e.g. in defending the first horn of her dilemma, she tells us:

> [T]he content of human evaluative judgements has been tremendously influenced...by the forces of natural selection, such that our system of evaluative judgements is saturated with evolutionary influence...[C]oincidence between the realist's independent evaluative truths and the evaluative directions in which natural selection tended to push us...would require a fluke of luck that's...extremely unlikely, in view of the huge universe of logically possible evaluative judgements and truths. (2006: pp. 121-122)

David Enoch makes clear that he thinks of the evolutionary challenge along those lines. According to Enoch, the challenge consists in the charge that a commitment to non-naturalist realism makes it apparently impossible to explain how our moral beliefs and the moral facts are reliably correlated, given that non-naturalistically construed moral facts are causally and constitutively independent of our moral attitudes. Here's Enoch's concise abstract statement of the challenge:

> Very often, when we accept a normative judgement j, it is indeed true that j; and very often when we do not accept a normative judgement j (or at least when we reject it), it is indeed false that j. So there is a correlation between (what the realist takes to be) the normative truths and our normative judgements. What explains this correlation? (Enoch 2011: p. 421)

Evolution highlights this problem. Assuming non-naturalist realism, the moral truths are not dependent on our moral attitudes. Furthermore, the moral non-naturalist is not a moral sceptic: she thinks that often enough, our moral beliefs land on the relevant moral truths, and not just as a matter of luck. Therefore, non-naturalists are committed to a reliable correlation between our moral beliefs and the moral truths.

But assuming the debunkers have gotten the empirical story right, our moral beliefs have been influenced to a significant degree by evolutionary pressures. Assume that the non-naturalist does not deny the empirical story. Now, as we have noted in section (1.4), the generic EDA tells us that non-naturalistically construed moral facts were not involved in the evolutionary explanation of our moral belief-forming faculty. Putting all this together, the non-naturalist moral realist now is committed to strong correlation between the independent moral truths and the moral beliefs you can expect in evolutionary successful creatures whose moral belief-forming tendencies have been selected for (Enoch 2010: pp. 425-426). But what accounts for this relation given that non-naturalistically construed moral facts are not involved in the evolutionary explanation of our moral faculty? What accounts for the strong correlation given that evolution is a truth-irrelevant influence, an influence that we should not expect to reliably push us towards the moral truths? So, according to Enoch (and it seems third-factor theorists more generally), the "general challenge…is that of coming up with an explanation of a correlation between our relevant beliefs and the relevant truths" (ibid.).

If this challenge cannot be met, it seems that the non-naturalist moral realist is committed to positing an unexplained brute correlation between moral beliefs and moral truths, i.e. a reliable correlation that came about by accident. And this seems to imply that whenever we hit on the truth in our moral beliefs, we do so as a matter of epistemic luck. This is due to (i) our beliefs being formed under the influence of a

truth-irrelevant force, and (ii) there being no principled explanation for the reliable correlation between our moral beliefs and the moral truths.

In turn, third-factor theorists try to neutralize this potential threat by providing an account of how we would end up with a reliable moral belief-forming faculty, *even though* non-naturalistically construed moral facts are not involved in the evolutionary explanation of the development of this faculty.

The guiding thought behind third-factor proposals is succinctly stated by Karl Schafer:

> [In] evaluating the reliability of our normative dispositions, it doesn't matter whether or not they developed so as to track the nonnormative properties that have normative significance *because* these properties have normative significance, so long as the development of our normative faculties was sensitive to the distinction between properties that do have normative significance and those that do not *for some reason*. (2010: p. 480; Schafer's emphasis)[71]

Let's illustrate Schafer's point by using the example of *pain*. Say that it is a moral fact that pain is *pro tanto* morally bad (I'll leave the "pro tanto" implicit in the next few lines). To meet the challenge of accounting for the reliable correlation between moral beliefs and truths, it is not necessary to claim that our moral faculty developed as it did because pain really is morally bad.[72] It will suffice to have an account that shows how our moral faculty has developed in a way that makes us disposed to believe that pain is bad for some reason. The reason here could be that as it happens, creatures who were disposed to believe (or proto-believe) that pain is bad in our ancestors' circumstances had a reproductive advantage over creatures who were disposed to believe (or proto-believe) otherwise. Since these beliefs happen to be true, it turns out

---

[71] Schafer's account here serves to emphasize a general thought underlying third-factor responses, even though it is certainly open to discussion, whether Schafer's account is best seen as a third-factor account, or as a closely related, but more straightforward response to EDAs (cf. Morton 2018: p. 4).

[72] This is good news too: this claim would tie you to the implausible position that Street termed the "tracking account" (cf. section (1.2.1). That means, the account that evolutionary forces have tended to make our evaluative judgments track the attitude-independent evaluative truths or facts because making true evaluative judgments promoted our ancestors' reproductive success

that evolution happened to lead to the development of a faculty that would dispose us to have true beliefs about the badness of pain.

Now, with some context in place, we are in position to look at third-factor accounts in some detail. I take the following succinct general statement of third-factor views from Selim Berker:

> ***THIRD-FACTOR ACCOUNT.*** Evolutionary forces have tended to make our moral judgments track the attitude-independent normative truth because, for each normative 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.[73] (2014: p. 15)

As Berker points out, one basic thought behind third factor-accounts is that it is not necessary to posit a direct dependency relation between our belief that p and the fact that p to account for why it is not accidental that the belief tracks the fact (Berker 2014: p. 15). In many cases, a common cause-structure or a structure akin to a common cause-explanation will serve just fine to explain the tracking relation between judgment and fact (where we posit a third factor on which the belief and the fact depend). Another point that is emphasized by Berker is that "when explaining why a given judgment tracks a given fact, any sort of a dependency relation is enough" (ibid.). So, to explain why a belief that p tracks the fact that p, one can take a common-cause structure and replace one of the causal relations with a metaphysical relation (e.g., a supervenience or a grounding relation) (ibid.). When "p" is normative we would then have a typical third factor-explanation, where you posit some non-normative factor F on which the belief that p causally depends, and on which the fact that p metaphysically depends. Importantly, this allows the realist to explain why our

---

[73] In the above "pain-example", the relevant factor F would be the non-normative facts about pain, which are then said to both (i) partly (causally) explain why creatures, who were disposed to believe (or proto-believe) that pain is bad in our ancestors' circumstances had a reproductive advantage, and (ii) to partly (metaphysically) explain why it is a normative fact that pain really is bad.

moral judgments track the moral facts, without endowing the moral facts with causal powers.

Different third factor accounts will substitute "F" with different non-normative assumptions, and they will then posit different substantive moral facts, that are (at least partly) grounded in F. For example, Enoch's proposal (2010, 2011) substitutes the "non-normative facts about survival (of creatures like us and our ancestors)" for "F", and works with the normative assumption that survival is good for beings like us and our ancestors. Wielenberg (2010) pursues a similar strategy in substituting "F" with the "non-normative facts about certain cognitive faculties (of creatures like us and our ancestors)" and he utilizes a moral assumption about the value of certain cognitive faculties.

To further illustrate how these accounts work, consider Enoch's and Wielenberg's proposals. As Enoch (2011) has argued, our moral beliefs about what is good track facts about what promotes survival – that's an important part of the evolutionary story about the origins of our moral beliefs. That survival indeed is generally good for beings like us in our usual circumstances is the substantive normative assumption involved.[74] As it happens, the argument goes, evolutionary forces have tended to push us towards believing that survival is good (because being disposed to believe this presumably tended to raise fitness in our ancestors' circumstances). Survival therefore stands both in a causal relation with our moral beliefs (as evolutionary forces have pushed us towards believing that survival is good) and it stands in a metaphysical relation with the moral facts (e.g., the moral fact that survival is good is partly grounded in the non-normative facts about survival or the moral fact that survival is good supervenes on the non-normative facts about survival). Thus, there is a third factor – survival – that explains both our moral beliefs and the moral facts. If survival is good for beings like us and evolutionary forces have pushed us towards moral judgments that track the goodness of survival, then it turns out that evolutionary forces have generally pushed us in the direction of the moral truth.

[74] The claim that "survival is good" is qualified in quite a few ways in Enoch's argument: he says that all he requires is that survival, for beings like us and our ancestors, in our usual circumstances, is by-and-large better than the alternative (Enoch 20111: p. 168).

Wielenberg (2010) assumes that creatures like us, who are endowed with certain cognitive capacities, have rights. Starting from this assumption, he argues to the conclusion that we are perfectly reliable at detecting that we have rights – and for evolutionary reasons, no less. Evolution has selected for creatures with our advanced cognitive capacities, once again presumably for the reason that having these capacities tended to raise fitness. Creatures with such advanced cognitive abilities have rights. The fact that humans have rights is partly grounded in the facts about their cognitive makeup. Thus, in selecting for creatures like us to have certain advanced cognitive capacities, evolution selects for creatures with rights. The crucial step for Wielenberg is to point out that to entertain the thought that you have rights, you must be a creature with advanced cognitive capacities that enable you to have some grasp on the concept of rights (2010: pp. 446-447). If rights exist, it is widely agreed their presence is guaranteed by the presence of advanced cognitive faculties (ibid.). Wielenberg assumes that humans have rights. Thus, if you believe that you have rights, you exhibit certain cognitive faculties which guarantee that you have rights (given that rights exist). It is thus no worrisome coincidence that our beliefs that humans have rights match the moral truth, as evolution pushed us towards developing cognitive faculties which also (partly) ground the fact that humans have rights.

Both Enoch and Wielenberg also discuss potential worries that their third-factor accounts might yet involve instances of problematic epistemic luck. The worry here might be that it is also a problematic instance of epistemic luck that evolutionary forces have played just the right sort of role in bringing about a correlation between our beliefs and the relevant truths, given that it is not the case that evolution has pushed us towards forming our beliefs because they are true. It seems that if our ancestors' circumstances had been relevantly different, evolution's influence on our moral faculty would have led to relevantly different (and thus false) beliefs. So, it seems that even given the third-factor explanation, there's a sense in which we are lucky to have true moral beliefs.

Enoch responds to this worry with what seems to be a "partners-in-guilt"-argument (Enoch 2011: pp. 172-173). He asserts that it is true that some luck is involved in evolving a faculty that happens to dispose us to believe truly. But he argues that the same kind of epistemic luck is involved in the development of accurate perceptual

faculties. Had our ancestors' circumstances been relevantly different, then being disposed to have accurate perceptual beliefs might not have raised fitness in those different circumstances. But we don't tend to worry about this kind of epistemic luck in the development of our perceptual faculties. Since, according to Enoch, it is the same kind of epistemic luck in both cases, either we should be worried in both cases or in neither of them. Since recommending being worried in both cases seems to settle you with a sceptical result that goes far beyond the debunkers' intended goal, Enoch concludes we should not worry in either case.

A second worry is that according to Enoch's proposal, evolution happened to aim at something which is good (i.e., survival) but might also have aimed at something that is not good.[75] And in that case, we would have tended to believe that this different aim is good and would thus have believed falsely. (You will recognize this thought from our above discussion on the issue of whether evolution easily could have resulted in us believing differently in moral matters.) Here Enoch seems to basically just state that for creatures like us and our ancestors in our circumstances, it does not seem to be the case that an aim different than survival could easily have led to belief-forming tendencies that would have been adaptive (cf. Enoch 2011: p. 172). Therefore, we need not worry about what would have been if evolution would have aimed at something different than survival.

One thing also worth pointing out is that third-factor accounts in themselves only offer a partial vindication of what we usually take ourselves to know in moral matters. They aim to show how certain basic individual claims to moral knowledge (concerning e.g. the goodness of survival or that humans have rights) are defensible even given an evolutionary moral genealogy. However, the hope of third-factor theorists clearly is that this can deliver the base for a more thorough vindication (Enoch 2011: p. 168), a vindication of many of our non-basic moral beliefs. The thought here seems to be that once we can take our basic moral beliefs for granted, then we can tell a story on how we can potentially use e.g. rational reflection to correct for the influence of

[75] A related worry could of course also be that survival could have failed to be good. But with regards to this worry, Enoch seems to go for the same reasoning as Clarke-Doane: that survival is good is a basic moral truth that is necessary in a strong sense (Enoch 2011: p. 172).

evolution, if we had reason to think that this influence would perhaps lead us astray in the case of some non-basic moral beliefs (Vavova 2015: p. 115). Basically, third-factor theorists like Enoch here seem to think that *if* the reliable correlation between our basic moral beliefs and the independent moral truths is not unexplainable, we will have some plausible means for accounting for the reliable correlation between our non-basic moral beliefs and the relevant independent moral truths as well (via e.g. pointing to the correcting influence of rational reflection).

Again, I think it is useful to end with putting the above material together in the form of an argument. This time around, the argument we end up with is a bit simpler:

THIRD-FACTOR ARGUMENT.

(i) Our moral beliefs are true.

(ii) Evolutionary considerations can only undercut our moral beliefs via giving us reason to think that the reliable correlation between our moral beliefs and the independent moral truths is unexplainable.

(iii) If our moral beliefs are true, then we can explain the reliable correlation between our basic moral beliefs and the independent moral truths via a third-factor account.

(iv) Therefore, the reliable correlation between our basic moral beliefs and the independent moral truths is not unexplainable.

(v) If the reliable correlation between our basic moral beliefs and the independent moral truths is not unexplainable, we can account for the reliable correlation between our non-basic moral beliefs and the relevant independent moral truths (via e.g. pointing to the correcting influence of rational reflection).

(vi) Therefore, the reliable correlation between our non-basic moral beliefs and the relevant independent moral truths is not unexplainable.

(vii) Therefore, evolutionary considerations do not undercut our moral beliefs.

That wraps up my presentation of the two standard responses to EDAs. The remainder of this thesis will be dedicated to the development of arguments against these two standard responses.

2.4 SUMMARY

In this chapter, I have presented two influential standard responses to EDAs, third-factor accounts and the AMS. These two responses offer considerations meant to neutralize the epistemic threat supposedly arising from EDAs. Both responses suppose that the epistemic threat in need of neutralizing is roughly that evolutionary considerations might show our moral beliefs to be subject to a problematic kind of epistemic luck. And both responses crucially assume the truth of our common-sense moral beliefs. In the next two chapters, I will now evaluate these responses.

3 THE STANDARD RESPONSES AS DEFEATER-DEFEATERS

3.1 INTRODUCTION

The generic EDA is meant to generate a defeater for our moral beliefs. Third-factor responses and the AMS try to establish the claim that evolutionary considerations do not undercut the justification of our moral beliefs. That means they both offer considerations meant to counter-act or neutralize the supposedly defeating force of evolutionary explanations. In this chapter, I try to do a few things. First, I want to elucidate what it is for a consideration to counter-act or neutralize a supposed defeater. Next, I will consider one possible option for how an alleged defeater can be neutralized (the defeater-defeater option). And I will assess the success of the standard responses, if we understand these responses as defeater-defeaters.[76]

Here is the plan for the chapter. First, in (3.2), I will work out how these two responses conflict with the generic EDA. Here I will also argue that in relation to the generic EDA (and EDAs more generally), we can either understand the AMS and third-factor accounts as presenting defeater-defeaters or defeater-deflectors.

In (3.3), I will show that for both third-factor accounts and the AMS, we can find cases, where, intuitively, a belief's justification is lost in virtue of incoming information about the explanatory history of the faculty that produced it. And I will argue that in those cases, intuitively, justification is lost, even though the believer is aware that if her belief were true, her belief is modally secure or not subject to a problematic sort of coincidence. I will furthermore argue that in all relevant respects, these cases are exactly analogous to the cases of the relevant moral believers. The two examples constructed in this section will be very important for the further critical discussion, in both this chapter and the next.

In (3.4), I will argue that if we understand the AMS and third-factor accounts as defeater-defeaters, we have good reason to think that both responses are not

[76] As will become apparent during this chapter, my line of reasoning draws heavily on Moon (2016).

successful in reinstating the justification of our moral beliefs. In (3.5), I will conclude by summarizing the most important points from this chapter.

3.2 DEFEATER-DEFEATERS OR DEFEATER-DEFLECTORS?

After the presentation of the argument from modal security and third-factor accounts in the previous chapter, I will now develop an argument against these standard responses to EDAs. Let's first work out where the epistemological conflict between these responses and the generic EDA lies.

To recap: How does evolution generate a defeater for your moral beliefs according to the generic EDA? According to the generic EDA, evolution generates an undercutting defeater for your moral beliefs by making it the case that you should withhold belief that the explanatory history of your moral faculty involves moral facts. If your belief that your moral faculty is reliable is not based on a source of justification independent of this faculty, then, in the end, this makes it the case that you should withhold or disbelieve that your moral faculty is reliable. This is the generic EDA's story about how receiving information that the best explanation for our moral faculty does not involve the relevant moral facts takes away your justification.

Both third-factor responses and the AMS argue for the claim that evolutionary considerations do not undercut the justification of our moral beliefs. That means they both offer considerations meant to counter-act or neutralize the supposedly defeating force of evolutionary explanations. In section (1.3.2) above, we heard that the epistemic danger arising from alleged defeaters can be met in two ways:

> **Via a *defeater-defeater***: Initially, at t_0, your belief that p is justified. Then, at t_1, you receive information D that defeats your belief that p. At this point in time, you are not justified to believe that p. But afterwards, at t_2, you receive new information D-DEFEATER that counter-acts the defeater for believing that p, which reinstates your justification for believing that p.
>
> **Via a *defeater-deflector***: Initially, at t_0, your belief that p is justified. At t_1, you already possess or receive information D-DEFLECTOR, that keeps information D, which you receive at the latter point in time t_2, from defeating your belief that p and thereby from taking away your justification for believing that p. Without

possessing or receiving D-DEFLECTOR at t_1, receiving information D at t_2 would have defeated your justification for believing that p.

The basic difference between defeater-defeaters and defeater-deflectors is a difference between how the justification for a belief of yours changes or remains constant depending on everything else you believe or should believe at certain points in time. In section (1.4.1), we formulated the following necessary condition for justified belief, a condition that basically tells you that for your belief to be justified, you need to lack an undercutting defeater for that belief:

> ***NO DEFEATER.*** S justifiably believes that p only if it is not the case that in virtue of S's doxastic/normative set of background beliefs, S should withhold or disbelieve p*.[77]

If S justifiably believes that p it is necessarily not the case that in virtue of her background beliefs or in virtue of beliefs that S should have due to the evidence available to her, S should withhold or disbelieve that her belief was reliably formed (via the method that S employed to form it).

In defeater-defeater-cases (where the relevant defeater is an undercutting one), the following scenario now obtains: at a certain point in time, t_0, S is justified in believing that p via method M. Afterwards, at t_1, due to new information that S receives, it is now the case that, given everything else she now believes or should believe, S should withhold or disbelieve that p*. And after that, at t_2, S comes into possession of further information, that counter-acts the information she received at t_2 in such a way, that S should now no longer withhold or disbelieve that p*.

Take the following example of such a scenario: Paula sees (or seems to see) a guitar in her brother's garden shed, and on the basis of her visual perception, she forms the belief: "There's a guitar in in my brother's garden shed". Initially, it seems that this belief formed on the basis of her visual perception is justified. Sometime afterwards, Dr. Uno, an ophthalmologist with seemingly excellent credentials, diagnoses Paula during a routine examination with a peculiar eye defect, an eye defect that makes her hallucinate guitars whenever she finds herself in a garden shed. It seems that receiving

---

[77] Where p* was the proposition that your belief that p was formed reliably (via your way of arriving at your belief that p).

this information indeed undercuts her original belief's justification. That means Paula's justification for believing that there is a guitar in her brother's garden shed is lost upon receiving the medical expert's testimony. Assume also that Paula has no reason to mistrust Dr. Uno. Later still, a few colleagues and associates of Dr. Uno come to Paula. While these people assert that there really is a peculiar eye defect that causes you to hallucinate guitars when you are in a garden shed, they also tell her that Uno recently went mad. This led him to diagnose people with exotic eye defects they do not have. Paula, it now seems, was one of those people. It seems that the justification for Paula's initial belief is reinstated again. Here the testimony of Uno's colleagues and associates counter-acts the defeating force of Uno's testimony in such a way as to make it the case that her belief that there is a guitar in her brother's garden shed is once again justified on the basis of her original visual perception.

In a defeater-deflector case (again, where the relevant defeater is an undercutting one), the scenario looks like this: at a certain point in time, t_0, S is justified in believing that p via M. Afterwards, at t_1, S is already in possession of information, or else now receives information, that serves to prevent the occurrence of defeat at t_2, where in the absence of this information, S's belief that p would have been defeated at t_2. That means that either before or at t_2, it is never the case that S's belief that p via M becomes unjustified.

Consider the following example of a defeater-deflector-scenario: Tim is a PhD-student, who is writing a thesis in history. He's working on the causes of an outbreak of a typhus epidemic in the city of ABC in the 19th century. Sometime into his investigation, after collecting and assessing considerable amounts of evidence, Tim forms the (let's suppose) initially justified belief that the outbreak was mainly due to mismanagement at the level of local government. He then discovers a fairly unknown study by another historian, Liz, that discusses a number of documents (let's call these documents "Dossier X") that Tim so far has not encountered in his research. Although Liz's study is unknown, it seems to be of a very high quality (which is attested by several experts on the subject at hand, which Tim presents with Liz's book), and makes a rather compelling case that is consistent with all the known facts about the case. Liz reports in her study that Dossier X looks very much authentic, and seems to indicate clearly that it was actually the national government's

mismanagement, which was to blame for the typhus outbreak. Dossier X also indicates that the national government then proceeded to cover its tracks via fabricating evidence that implicates the local government. And it is this seemingly fabricated evidence, on which Tim has based his originally justified belief.

But Liz then goes on to argue, convincingly, that the whole case (supported by Dossier X) for a conspiracy headed by the national government is made up. Through painstaking research, she has discovered that Dossier X, which seems to show that there was a conspiracy by the national government to avert blame for the outbreak of the typhus epidemic, was most likely itself manufactured by a masterful forger in the service of the local government of ABC. The forged Dossier X is so convincing, that it took Liz years of work to prove that it is not authentic. Liz concludes that the (most likely forged) Dossier X gives us no reason to think that there's anything wrong with the original evidence (=the evidence which indicates that the local government of ABC is responsible).

Sometime afterwards, Tim comes across the fabricated documents that make up Dossier X. Had he not discovered Liz's almost forgotten study, then Tim's belief that the local government was crucially responsible for the outbreak would have been defeated via casting doubt on the trustworthiness of the grounds for Tim's original belief. But the information contained in Liz's study seems to be capable of preventing the occurrence of defeat (via preventing Dossier X from functioning as a defeater). Thus, in this example, it is never the case that Tim's initially justified belief becomes unjustified.

The generic EDA tries to generate an undercutting defeater for our moral beliefs. In response, third-factor accounts and the AMS can be seen as responding either via presenting a defeater-defeater or a defeater-deflector. That means these responses can either argue that our justification for our moral beliefs can be reinstated (the defeater-defeater option) or they can argue that the justification for our moral beliefs is never lost to begin with, since the considerations offered by these responses prevent the evolutionary explanation from constituting a defeater in the first place (the defeater-deflector option).

So, the question I am interested in now is:

Do third-factor responses and the AMS constitute defeater-defeaters or defeater-deflectors?

I will go through both options in this chapter (the defeater-defeater option) and the next (the defeater-deflector option). The trouble for the AMS and third-factor accounts is that neither option is promising.

3.3 A Set of Inconvenient Cases: Larry and Anne

In this section, I will argue that for both third-factor accounts and the AMS, we can find cases where, intuitively, a belief's justification is lost in virtue of incoming information about the explanatory history of the faculty that produced it. And I am going to argue that in those cases, intuitively, justification is lost, even though the believer is aware that if her belief were true, her belief is modally secure or not subject to a problematic sort of coincidence. I will furthermore argue that in all relevant respects, these cases are exactly analogous to the cases of the relevant moral believers. There are ways to dispute the analogy, but these ways of responding seem to make the standard responses superfluous.

As we are going to see in the next two sections, the existence of these two kinds of cases proves to be very inconvenient for the AMS and third-factor account. Intuitively, in both cases, the relevant subjects would be epistemically irrational for continuing to hold onto their belief in light of the relevant incoming information.[78]

3.3.1 The Case of Larry

Let's begin by discussing an example that is problematic for the AMS. Here I present and discuss a version of an example due to Silvia Jonas (2016).[79] Larry believes that

---

[78] My presentation of the cases in (4.3.1) and (4.3.2), and the subsequent discussion of the standard responses as defeater-defeaters in (4.4) is similar in structure to Locke's (2014: p. 231) discussion of his "Martian"-case.

[79] Jonas uses the original version of this example to argue that the AMS is problematic for rendering certain beliefs "viciously immune" (cf. Jonas 2016: p. 11; Schechter 2018). Jonas argues that if sound, Clarke-Doane's argument would make it in principle impossible to challenge the reliability of beliefs that have as their content metaphysically necessary truths and with respect to which some intervening force (e.g., evolution or cultural influence) has made it so that we could not easily have believed differently. However, both Jonas and Schechter make the point that this is highly implausible: sometimes causal influences in the origin story of these beliefs do seem to pose a real challenge to the reliability of these beliefs (Jonas 2016: p. 11).

God exists. This belief is in fact true. This belief is based on an intellectual seeming[80] of Larry. It seems to Larry that God exists, and, on this basis, he forms the belief that God exists. Assume that Larry's belief that God exists is initially justified. Significantly for our purposes, this means (among other things) that given his initial doxastic/normative set of background beliefs, it is not the case that Larry should withhold or disbelieve that his belief that God exists was formed reliably (via an intellectual seeming).

Note that Larry's theistic belief has important similarities with the moral beliefs of ordinary epistemic agents. First, many of the moral beliefs of ordinary believers seem to be based on their strong inclination to have these beliefs. In other words, many ordinary beliefs are held by people because they seem true to them. For example, it appears that ordinary moral believers often think that killing is (pro tanto) wrong because it seems true to them that killing is (pro tanto) wrong, and this seeming remains intact even after some reflection.

Now, perhaps Clarke-Doane would object at this point by saying that the crucial difference between Larry's theistic beliefs and the beliefs of ordinary moral believers is that ordinary moral beliefs are initially justified, while Larry's belief that God exists is not. But it is hard to see why that must be case. Stipulate that Larry has thought as long and hard about God's existence as any typical moral believer has thought about the wrongness of killing or the value of giving to charity. Stipulate that Larry has so far not encountered evidence that indicates clearly that there might be something epistemically wrong with his belief. My point here is basically the following: if it can

---

[80] Intellectual seemings are *sui generis* propositional attitudes that p, akin to perceptual seemings, that can non-inferentially justify beliefs formed on their basis. What distinguishes seemings from other attitudes is their peculiar phenomenal character:
"The phenomenology of a seeming makes it feel as though the seeming is 'recommending' its propositional content as true or 'assuring' us of the content's truth." (Tucker 2011: p. 57)
I basically assume that *intuitions* are intellectual seemings here. If you are uncomfortable with an account of intuition in terms of intellectual seemings, you could substitute it with an account of intuitions as e.g., non-inferentially justified beliefs.
I grant that, in virtue of having the intuition that p, I have *prima facie* justification to believe p, even if I have no positive evidence for the reliability of my intuitions. As Mogensen (2017: pp. 282-283) writes, this is something the debunker should grant: given the extent to which our moral judgments are governed by our intuitions, supposing otherwise would arguably require a wide-ranging scepticism about ordinary moral beliefs, rendering any concern about the debunking power of evolutionary considerations idle. For more on intellectual seemings, see FN 153 in appendix I.

be the case that (in the right circumstances) moral beliefs gain some modest level of epistemic support via being based on a subject's intellectual seeming, then it is hard to see how one could exclude the possibility that Larry's theistic belief (in the right circumstances) gains the same level of support from being based on his intellectual seeming. I simply ask you to imagine that the right circumstances are in place: i.e., it is initially at least not the case that given everything else Larry believes or should believe, Larry should immediately withhold or disbelieve that his belief that God exists was reliably formed.[81]

Let us furthermore stipulate that Larry is aware and justifiably believes that if his belief that God exists is true, it is metaphysically necessarily true.[82] Now, given that I suppose that Larry is an ordinary theistic believer in this scenario, he does not, of course, explicitly believe "If my belief that God exists is true, it is metaphysically necessarily true". But Larry could well be aware that if God exists, then there is a certain sense in which God could not easily have failed to exist. And assume that if he were to be interrogated by a philosopher, Larry perhaps could be brought to agree and appreciate that there is a clear difference in necessity between the existence of God (assuming God exists) and the existence of e.g., a book in Larry's library. And perhaps he could, on the basis of this appreciation, be brought to assert that if his belief that God exists is true, it is necessary in the same strong sense as his belief that atoms of gold have 79 protons is necessarily true. And that suffices for the purposes of the example.

Moreover, let's say that Larry is aware and justifiably believes that given his circumstances, he could not easily have believed differently. In other words, he justifiably believes in a certain explanation for why he holds the belief that God exists,

---

[81] It is also important to note that the requirements on Larry's circumstances do not seem to be that challenging, at least if we are assuming that the moral beliefs of actual moral believers can be and regularly are justified on the basis of their moral intuition. If a substantial number of actual moral believers hold moral beliefs, which are justified on the basis of their intuition, then it must be the case that the requirements on a believer's circumstances allow for factors like e.g., the awareness, that there is substantial disagreement about the subject matter of the relevant beliefs.

[82] I am assuming that Larry's theistic belief is about the God of "Perfect Being Theology", i.e. an omnipotent, morally perfect and omniscient divine being that is often thought to be metaphysically necessary, if it exists (cf. Davidson 2013: Section 1).

an explanation that has no bearing on the truth of the belief,[83] but that does show that Larry could not have easily believed differently. So, Larry in this example is broadly aware and justifiably believes that his belief that God exists, if true, is metaphysically necessary, and that given his circumstances, he could not easily have believed differently. Thus, I stipulate that it is correct to say of this scenario that Larry is broadly aware and justifiably believes that his belief that God exists is modally secure *if it is true*.

So, Larry is broadly aware and justifiably believes that if his belief that God exists is true, then it was all but bound to be true. However, Larry then receives information that for the whole period of time in which he has formed and held the belief that God exists, he has been fed hallucinogenic drugs. And he's informed that these are drugs, which tend to produce intellectual seemings that God exists in most people.

One plausible way of working out what happens when Larry receives this information is to say that he has now come into the possession of evidence for the claim that the best explanation of the faculty that produces his intellectual seemings about theistic subject matters does not involve facts about God. This makes it the case that Larry should now withhold belief about the claim that the explanatory history of his "theistic intellectual seemings"-faculty involves the relevant theistic facts.

Now, Larry might not be too troubled by this. After all, if Larry has means to confirm the reliability of his "theistic intellectual seemings"-faculty which are independent of this faculty, then it seems that his belief might nonetheless be justified, even after Larry has considered all the information now accessible to him. But let us stipulate that Larry does not have independent means of doing this. That means Larry has no means of confirming the reliability of his faculty that produces his theistic intellectual seemings that is independent of this very faculty.

So, in line with the general epistemological story that supports the generic EDA, we can give an account of how receiving information that his theistic beliefs have been formed under the influence of hallucinogenic drugs affects the initially positive epistemic status of those beliefs. Given all the information in his possession, it is now

---

[83] Knowing that this explanation holds does not give Larry an epistemic reason to think that his belief is true.

the case that Larry should withhold or disbelieve that his theistic beliefs have been formed reliably via his "theistic intellectual seemings"-faculty. And this means that his justification for holding his theistic beliefs is undercut.

Importantly, this epistemological story stays silent on the modal security of Larry's belief. Since Larry is broadly aware that if his belief that God exists is true, this belief is modally secure, he does not think that the above incoming information makes his belief that God exists epistemically inappropriate *because* it shows it to be not modally secure.

Is the case of Larry exactly analogous in all relevant respects to the case of the moral believer according to the AMS? It seems it is. In both cases, the relevant beliefs are assumed to be true. In both cases, the relevant beliefs are initially justified. In both cases, these beliefs are modally secure, if true, and the subject is aware that her beliefs are modally secure, if true. In both cases, the subject is confronted with information that makes the subject aware that the best explanation for her relevant way of arriving at her beliefs does not involve the relevant facts. And the subject becomes aware that she lacks independent means of confirming the reliability of her way of arriving at her beliefs. So, with respect to all these features, the two cases seem to be exactly alike.

3.3.2 THE CASE OF ANNE

Let's now provide an example that is relevant to the discussion of third-factor accounts. Here I use a slightly amended example due to Lutz (2017: p. 17) (who takes this example from (Bedke 2009)). Consider the case of Anne.

Anne believes that every person has a spirit animal that is connected to her personality. Wise people have owls, brave people have lions, and so on. Anne forms these beliefs based on intellectual seemings. For example, when she interacts with Jordan, who is wise, she has an intellectual seeming that Jordan has an owl as her spirit animal, when she interacts with brave Grigori, it seems that her that Grigori has as his spirit animal a lion, and so on. Based on her seemings, she goes on to form beliefs like "Jordan has an owl as her spirit animal" and "Grigori has a lion as his spirit animal". Assume that Anne's beliefs about spirit animals are actually true.

Let's say that, initially, Anne's beliefs about spirit animals are justified. Among other things, this means that, initially, it is not the case that in virtue of her

doxastic/normative set of background beliefs, Anne should withhold or disbelieve that her beliefs about spirit animals have been formed in a reliable fashion via her relevant intellectual seemings. Perhaps third-factor theorists like Enoch and Wielenberg would now object that the case of Anne and the case of the typical moral believer are disanalagous precisely because Anne's beliefs are not initially justified. Third-factor theorists could simply state that in contrast to the moral convictions of ordinary moral believers, Anne's spirit animal-beliefs are not initially justified since they are just so outlandish that it is hard to see how it could be epistemically sound for Anne to trust her intellectual seeming.

As I have said above, S's intellectual seeming that p can only render a belief that p justified in the right kind of circumstances, i.e. in circumstances in which it is not the case that S immediately has sufficient reason to think that this intellectual seeming that p is not trustworthy or that it is not a good indicator of the truth about p. Against the claim that the sheer outlandishness of Anne's belief disqualifies it, I hold that there is no *in principle* reason to think that Anne's initial circumstances could not be such that they could render her spirit animal-beliefs based on her seemings initially justified.[84] Stipulate that Anne initially does not already have evidence or information that is sufficient for making it the case that she should withhold or disbelieve that her belief that Jordan's spirit animal is an owl is reliably formed by her "spirit animal intellectual seeming"- faculty". For example, assume that in the circumstances in which Anne initially finds herself in, there's no strong reason accessible to Anne to dismiss her spirit animal-seemings out of hand. Given this stipulation, I see no *in principle* reason for why Anne's beliefs could not enjoy some initial justification based on her intellectual seemings.

But now Anne receives information that what causes her intellectual seemings about spirit animals is a brain tumour. In other words, Anne has these seemings and her subsequent beliefs not because people do have spirit animals, but because of the tumour. Again, we can account for the effect that receiving this information has on

[84] Indeed, it seems that in many societies throughout history, a belief of this sort might have been perfectly acceptable, or at least as acceptable as moral convictions based on moral intuitions in our circumstances.

the initially positive epistemic status of her beliefs in terms of the epistemological story developed in the last chapter.

In virtue of the information about the tumour, it is now the case that the best explanation available to Anne for her "spirit animal intellectual seeming"- faculty does not involve the relevant facts about spirit animals. In line with the epistemological story developed in (1.4), it is now the case that Anne should withhold belief about the claim that the explanatory history of her "spirit animal intellectual seeming"- faculty involves the relevant facts. Suppose further that Anne lacks independent means of confirming the reliability of her way of arriving at those beliefs. At the very least, it appears that Anne would be epistemically irrational if she did not withhold or disbelieve that her belief that Jordan's spirit animal is an owl is reliably formed by her "spirit animal intellectual seeming"- faculty.

It is important to note that this information about what best explains her intellectual seemings is silent on whether some explanation is available to Anne, that shows how her beliefs and the relevant truths could be reliably correlated. For all that this account of how Anne's belief is undercut tells us, Anne might yet be able to reason in the following way:

> The discovery that that I have a brain tumour, which is responsible for my beliefs about spirit animals, does not ultimately defeat my belief. I have not learned anything, which is, strictly speaking, inconsistent with the existence of spirit animals. In the context of this sceptical challenge I am entitled to make some assumptions about the way the world is. Therefore, I will assume that all my spirit-animal beliefs are true. Since my beliefs about spirit animals are true, I guess the brain tumour made me reliable!

Moreover, assume that Anne has an explanation available for her reliability:

> **ANNE'S THIRD-FACTOR**. There is a factor, namely the factor that Jordan is wise, of which it is the case that:
> (i) Jordan's being wise is part of what explains why I think her spirit animal is an owl (i.e., Jordan's being wise causally helps make it the case that it seems to Anne that Jordan's spirit animal is an owl),

(ii) and her being wise is also part of what explains why her spirit animal is an owl (i.e., Jordan's being wise metaphysically helps make it the case that her spirit animal is an owl, since (let's assume) the non-natural facts about spirit animals strongly supervene on relevant psychological facts).

In other words, Anne now believes that, if Jordan's spirit animal is an owl, then she can provide a story about how it is not just a coincidence that she believes Jordan's spirit animal to be an owl.[85] And this gives Anne an account of how her intuitively true beliefs are produced in a way that is not epistemically lucky in a problematic sort of way: if Anne's belief that Jordan's spirit animal is an owl is true, then she has an explanation of the reliable correlation between her belief and the truth in hand. So, the discovery that she has a brain tumour does not show that she has been lucky in a problematic sort of way in landing on the truth, if her belief that Jordan's spirit animal is an owl is true. And Anne can then conclude her reasoning by stating: "And my belief that Jordan's spirit animal is an owl is true! Therefore, my justification for holding this belief is reinstated."

Once again, with respect to all relevant features, the case of Anne, and the case of the moral believer according to the third-factor accounts seem to be exactly analogous. In both cases, we assume that the relevant beliefs are true. In both cases, the relevant beliefs are initially justified. In both cases, the subject is confronted with information that makes the subject aware that the best explanation for her relevant way of arriving at her beliefs does not involve the relevant facts. And the subject becomes aware that

---

[85] Perhaps someone could challenge Anne's third-factor explanation in the following way: even if it is true that Jordan's spirit animal is an owl, it could easily have been the case that Jordan's spirit animal is not an owl, while Anne would still believe that Jordan's spirit animal is an owl, since Jordan would still have been wise. Taking a page from Bedke's original presentation of the example, I would like to point out that Anne could go for her own version of a strong supervenience-argument to defend the necessity of the claim that wise people have owls as spirit animals:
"When asked about other possible worlds, and other possible people, it is clear that Andy [the subject corresponding to Anne in Bedke's presentation of the example] believes that which spirit animal a person has strongly supervenes on that person's psychology (character traits, really). People who are predominantly brave have lion spirit animals; people who are predominately wise have owl spirit animals; etc. Andy acknowledges that these subvenient psychological properties are fully natural. Moreover, he thinks that the truths about spirit animals satisfy strong global supervenience: any two metaphysically possible worlds that differ with respect to who has what spirit animal would also differ with respect to natural properties." (Bedke 2009: p. 197)

she lacks independent means of confirming the reliability of her way of arriving at her beliefs. What's more, in both cases, the subject has a third-factor explanation available to her, that can account for the correlation between her beliefs and the relevant truths. Again, with respect to all these features, the two cases seem to be exactly alike.[86]

3.4 THE DEFEATER-DEFEATER OPTION

In this section, I will develop an argument to the conclusion that if we understand the AMS and third-factor accounts as defeater-defeaters, then there is good reason to think that both responses are ineffective. That means they are not successful in defeating the defeater generated by the generic EDA, and so they are not successful in reinstating the justification for our moral beliefs.

My argument will be rather straightforward. With the examples of Larry and Anne, for both third-factor accounts and the AMS we have found cases, analogous to the cases of the relevant moral believers, where, intuitively, a belief's justification is lost in virtue of incoming information about the explanatory history of the faculty that produced it. In those cases, intuitively, justification is lost, despite the fact that the believer is aware that if her belief were true, her belief is modally secure or not subject to an epistemically problematic sort of coincidence. What accounts for the fact that, in those cases, justification cannot be reinstated via the kind of reasoning suggested

---

[86] The cases of Larry and Anne will be very important to the argument in the rest of this chapter, and to the argument in the next chapter. I have argued that with respect to all relevant features, the cases of Larry and Anne are exactly analogous to the cases of the relevant moral believers. But let me clarify that I do think that certain philosophers, who are sceptical of EDAs, have the resources to deny the analogy. The (to my mind) most straightforward way of denying the analogy is to argue that the empirical case of the debunker does not hold water, and that therefore an argument like the generic EDA does not generate a defeater at all, and therefore, does not generate a defeater in a way analogous to the defeater generated in the cases of Larry and Anne. I think that this might be a promising thought to pursue for the opponent of evolutionary debunking. Unfortunately, this line of argument also threatens to make the standard responses superfluous.
If you argue that the empirical case that the debunker needs to make is not sufficiently supported by the evidence or rests on confusions regarding evolutionary explanations, then it seems that there is no incentive to engage in the kinds of arguments proposed by Clarke-Doane and the third-factor theorists in the first place. Therefore, it seems that disputing the empirical case that supports arguments like the generic EDA is incompatible with thinking that the standard responses are non-superfluous to the debate. This is bad news to the standard responses, though obviously not necessarily good news for the debunker, as the actual empirical case of the debunkers looks vulnerable to empirically-minded replies, cf. appendix C.

by the AMS and third-factor accounts, is that this reasoning would be epistemically circular in a problematic kind of way. The examples I have presented in the last section will also be important in chapter (4), since there too I rely on the claim that these examples intuitively show at least that it would not be epistemically rational for the relevant subjects to continue to hold their respective beliefs.

As I have written above, both the third-factor accounts and the AMS argue that if our relevant beliefs are true, and if certain other conditions are in place, then these beliefs can be shown to be reliably formed. Without assuming that the relevant beliefs are true, neither the AMS nor third-factor accounts could get off the ground. To protect our common-sense moral beliefs, both responses therefore need to assume the following claim:

> **COMMON-SENSE MORALITY**. Our common-sense moral beliefs are by-and-large true.[87]

The central point in this section will be that, intuitively, in the presence of a defeater, a believer cannot reinstate her justification for her relevant belief via continuing to assume that her belief is true. The discussion of whether the third-factor theorist is entitled to COMMON-SENSE MORALITY is the focus of much attention in the current debate on EDAs (cf. Bedke 2009; Crow 2016; Locke 2014; Lutz 2017; Moon 2016; Vavova 2016; Wielenberg 2016). This point has not been as prominent in the critical discussion of the AMS, although the AMS too depends on this assumption (but cf. Morton 2018). My argument in this section aims to show that it is clearly not legitimate to rely on *prima facie* defeated beliefs to reinstate the justification of those very beliefs. If the AMS and third-factor accounts are meant to serve as defeater-defeaters, this spells serious trouble for them.

From here on out, until the end of this chapter, I will assume that we ought to understand the AMS and third-factor explanations as defeater-defeaters. If we understand the AMS and third-factor accounts as defeater-defeaters, then we have the following initial scenario before us with regards to both responses. After being confronted with information that makes it the case that you should withhold belief

---

[87] Both responses need to feature an assumption of this approximate scope to afford us a wholesale defence against EDAs.

about the claim that the explanatory history of your moral faculty involves moral facts, and since you are unable to independently confirm the reliability of your moral beliefs, the justification of your moral beliefs is undercut. That means it is now the case, that given everything else you believe or should believe, you should withhold or disbelieve that your moral beliefs are reliably formed.

Now assume that third-factor accounts and the AMS try to reinstate the justification of our moral beliefs via defeating the defeater. In other words, they are meant to provide considerations that make it the case that despite the presence of the evolutionary explanation for your moral faculty, and even though you lack independent means of confirming the reliability of your moral faculty, it is no longer true that you should withhold or disbelieve that your moral beliefs are reliably formed by your moral faculty.

The AMS aims to defeat the defeater via showing that, if your relevant beliefs are true, and if certain other conditions are in place, then your beliefs can be shown to be modally secure. Believing that *if* your belief is true, then it is modally secure, and assuming that your relevant beliefs are actually true, then, is meant to make it the case that you no longer should withhold or disbelieve that your moral beliefs are reliable. In a similar vein, third-factor accounts aim to defeat the defeater via showing that, if your relevant beliefs are true, then we have an explanation for how they are reliably correlated with the truth. Believing that *if* your belief is true, then it is reliably correlated with the truth, and assuming that your belief is actually true, then, is meant to make it the case that you no longer should withhold or disbelieve that your moral beliefs are reliable.

And here is how my argument in this section is straightforward: I am going to use the analogous cases of Larry and Anne to argue that in neither case is it legitimate for the relevant moral believer to rely on the truth of her *defeated* beliefs.[88]

Before I develop this argument, let me preclude potential confusion by addressing a possible objection. Here's the objection:

---

[88] My argument here is in line with Moon's (2016).

> An undercutting defeater for your belief that p does not provide you with a reason for believing that your belief that p is false. Since EDAs try to generate undercutting defeaters, is it not the case that the debunker, in the context of this sceptical challenge to the non-naturalist realist, must grant the realist the assumption that our moral beliefs are by-and-large true? If this is indeed so, how can it be epistemically illegitimate to rely on the truth of one's moral belief in neutralizing the epistemic threat supposedly arising from evolution?

Let me get this clear: an undercutting defeater indeed does not and cannot show that the relevant belief is false. So, the debunker must concede that evidence about the evolutionary influence on our moral beliefs does not entail the falsity of those beliefs. But that does not mean, as I will show in a moment, that it is legitimate for any believer to continue to rely on a set of beliefs in the presence of an undercutting defeater for those very beliefs. Admitting that evolutionary considerations do not and could not entail the falsity of our moral beliefs does not entail that we remain free to assume that our moral beliefs are by-and-large true after we have been confronted with these considerations. In short: conceding that a piece of evidence or information does not entail the falsity of your belief that p does not entail the further concession that, after you have been presented with this piece of evidence or information, you remain free to assume that your belief that p is in fact true.

Let us start with a general observation about how responses to undercutting defeaters can fail. In general, it is not legitimate to rely on the truth of defeated beliefs to show that your belief is reliable. To see this, consider the following example (which is a variation of an example due to Moon (2016: p. 13)):

> ***XX DEFEATER***. You ingest a pill called XX. You have no information about any negative effects of XX. You go outside to the driveway, and you see a red car parked there. Plausibly, your perceptual belief that there is a red car in front of you is defeasibly justified. But then you learn that XX makes 95% of those who ingest it see red cars even when there are no red cars in front of them. Here, you have a defeater for your belief via perception that there is a red car in front of you.

It is clearly illicit to engage in the following kind of reasoning in trying to defeat the relevant defeater:

> I have learned that XX makes 95% of those who ingest it see red cars even when there are no red cars in front of them. But I can see that there is a red car in front of me. So, as it happens, I must belong to the lucky 5%!

Analogously, after you have received information that your moral faculty is untrustworthy, it also seems illicit to rely on the truth of the beliefs produced by this very faculty to reason your way to the conclusion that your moral faculty really is reliable.

But this general point does not suffice to get us to the conclusion that, in the specific cases where the believer has further information available to her (e.g., that if her belief is true, it is modally secure), it is also true that it would be illegitimate to rely on the truth of a belief in the presence of a defeater for that belief. It does not suffice since the case of the subject in XX DEFEATER is not exactly analogous with the relevant cases of the moral believers according to the AMS and third-factor accounts. For instance, even if your belief that there is a red car in front of you is true, it is certainly not metaphysically necessarily true.

Luckily, as I have tried to show above, there are cases that are exactly analogous to the cases of the relevant moral believers, and in those cases it seems that the incoming information about what best explains the subjects' beliefs has epistemic consequences. Let us discuss the case of Larry first.

Intuitively, it seems clear that:

(i) Larry may well take his belief to be epistemically inappropriate, i.e. he may withhold belief about the claim that his theistic belief was reliably produced given that his belief has been formed and held by him under the influence of hallucinogenic drugs.

(ii) Moreover, it also seems that if Larry then takes his belief that God exists to be epistemically inappropriate given all the information now in his possession, this response is the epistemically rational response to his situation, given everything else that he has come to believe or should have come to believe about his circumstances. From the vantage point of what Larry believes or should believe about his situation, it seems it would only be epistemically

> rational for him to take a negative perspective towards his way of arriving at the belief that God exists.

This raises the question: Why does this response seem to be the only epistemically rational one, given that Larry is aware and justifiably believes that if his belief is true, it was all but bound to be true? This question is pertinent, as given what we have said about this scenario so far, Larry could now reason in the following way:

> Sure, no facts about God were involved in the best explanation for my way of arriving at beliefs about God. But my belief that God exists is true! What's more, I can see that given that it is true, it could not have been easily false, and furthermore, I could not easily have failed to have this belief. So, I was all but bound to get it right, and the presence of this explanation, and the lack of independent confirmation do not change that.

I take it as intuitively obvious that something is wrong with this kind of reasoning. Yet it is precisely the kind of reasoning that the AMS urges on us. But how is it wrong? The epistemic wrongness that is exhibited by Larry's reasoning here is the same kind of epistemic wrongness involved in the reasoning of the subject in the XX DEFEATER-case. To get this kind of epistemic wrong into view, we now need to say a few things about *epistemic circularity*.

Moon (2016) argues convincingly that the debate on what is admissible as a defeater-defeater is tied up with the debate on when epistemic circularity is problematic and when it is not. For example, my belief that my memory is overall reliable is epistemically circular if, for its justification, it depends itself on my memory (ibid.: p. 9). As Moon points out, many epistemologists think that epistemic circularity need not always be problematic, as these epistemologists argue that "a belief's being epistemically circular does not alone disqualify it from being justified; sometimes it does disqualify it, sometimes it doesn't" (ibid.: pp. 9-10). There are malignant and benign cases of epistemic circularity, where malignant epistemic circularity disqualifies a belief from being justified, while benign circularity does not.

Michael Bergmann (2006: pp. 198-200), who has done influential work on epistemic circularity, argues for the following sufficient condition for malignant circularity. For a subject S, her belief-forming faculty M, and her belief that p produced by M:

> If S already has a defeater for believing that p is reliably produced by M, then using her belief that p, which is a deliverance of M, as her support for continuing to believe that M is reliable results in malignant epistemic circularity. (ibid.; Moon 2016: p. 10)

As Moon comments:

> [i]f one already has a reason to doubt source [M]..., then one cannot bootstrap one's way out of doubting source [M]… by relying on the deliverances of [M]… (2016: p. 10).

In a case of malignant epistemic circularity, a subject seriously questions or doubts, or should seriously question or doubt, the trustworthiness or reliability of her way of arriving at her relevant beliefs to the point where she should withhold or disbelieve that her beliefs are reliably formed in this way. In this situation, any further dependence on the subject's part on her original way of arriving at her relevant beliefs results in malignant epistemic circularity and is thus unfit to reinstate the justification for those very beliefs.

In his reasoning, Larry relies on his belief that God exists (produced by his theistic intellectual seemings-faculty), to show how his theistic intellectual seemings-faculty is a reliable way of arriving at beliefs about God's existence. So, this belief is epistemically circular. If Larry is a subject in a defeater-defeater-case, then, *ex hypothesi*, we are assuming that his initially justified belief that God exists, formed via his intellectual seemings, is undercut when he receives information about what best explains his theistic intellectual seemings. Given Bergmann's sufficient condition for malignant circularity, Larry's belief that God exists is therefore malignantly circular. Relying on a malignantly epistemically circular belief cannot reinstate your justification for holding this very belief. The justification for Larry's belief that God exists is therefore not reinstated via his ability to reason in the above way.

With Larry, we have a case of a subject whose belief is modally secure if true, and who also justifiably believes that his belief is modally secure, if true. Nonetheless, it would be epistemically wrong for Larry to try to reinstate his justification in the face of the above incoming information by assuming that his belief that God exists is true. It seems that Larry does not have a suitable defeater-defeater at his disposal.

What's more, the case of Larry seems exactly analogous to the relevant case of a moral believer according to the AMS. This is a moral believer who is confronted with information pertaining to the influence of evolution on her moral beliefs. Evolution makes it the case that the best explanation available to this moral believer for her moral faculty does not involve moral facts, and this moral believer lacks independent means of confirming the reliability of her moral faculty.

Now, our question was: if we assume that

(i) the moral beliefs of this subject are true,
(ii) that if her beliefs are true, they are modally secure,
(iii) and if we assume that this subject justifiably believes that if her beliefs are true, they are modally secure,

does this suffice to reinstate the justification of her moral beliefs?

The case of Larry seems to be exactly analogous with respect to all these features. In the case of Larry, I have argued that we should answer the corresponding question negatively. Given that the two cases are exactly analogous with respect to all the relevant features, I contend that the answer to this question should therefore also be negative.

In a moment I will say something about my assumption in the discussion of this case (and in the discussion of the next case) that malignant epistemic circularity disqualifies a belief as a suitable defeater-defeater.

But let's first discuss Anne's case. Once again, I regard it as intuitively obvious that Anne's reasoning does not reinstate her justification for believing that Jordan's spirit animal is an owl. But this is precisely the kind of reasoning that the third-factor theorist urges on us. Given all the information in her possession, it is an epistemically inappropriate response for Anne to continue to hold her belief that Jordan's spirit animal is an owl.

My explanation for this is again basically the same as in the case of Larry. Even if Anne's belief is true, and even if ANNE'S THIRD FACTOR shows that if her belief is true, then Anne's belief is not true as a matter of epistemic luck, Anne's response

to her situation is epistemically wrong for involving a malignant kind of epistemic circularity.

With the case of Anne, we have an example of a subject for whose belief there is a third-factor explanation, if this belief is true, and where this subject also believes that if her belief is true, then she can provide a story about how it is not just a coincidence that she believes truly.

Yet again, it would be epistemically wrong for Anne to rely on the truth of her belief in the face of defeating evidence for this very belief. Assuming that Anne is in a defeater-defeater-case is tantamount to saying that being informed that her belief that Jordan has as a spirit animal an owl is best explained by her tumour (and that she is unable to independently confirm the reliability of her spirit animal-intellectual seeming-faculty) takes away her justification for believing this. But once the justification for this belief is lost, it becomes epistemically inappropriate to further rely on the truth of this belief in her reasoning.

Furthermore, the case of Anne again seems to be exactly analogous to the case of the relevant moral believer according to third-factor accounts, who is faced with an undercutting defeater generated by the generic EDA.

I hope what I have said so far in this section suffices to establish the claims that (i) as instances of defeater-defeaters, both third-factor accounts and the AMS are clearly problematic, and (ii) they are problematic in a similar kind of way.

My line of reasoning so far plainly assumes that involving malignant circularity disqualifies a response from being a successful defeater-defeater. But this assumption (shared by Bergmann and Moon) seems very plausible. As the cases of Larry and Anne demonstrate, there are many cases, in which a subject could show that her belief is indeed reliably formed if she were to assume that her belief is true. Assuming that this move is permissible would serve to immunize beliefs from counter-evidence, when, intuitively, this counter-evidence still negatively affects the justification of those beliefs, as in the above cases. Therefore, declaring epistemically malignantly circular responses to be successful defeater-defeaters seems to conflict with our intuitive application of the notion of undercutting defeat. This, I take it, is why some philosophers working on EDAs have suggested that assuming that it is epistemically

unproblematic to rely on a set of beliefs in the presence of an undercutting defeater for those very beliefs threatens the "very intelligibility of the notion of undercutting defeat" (Lutz 2017: p. 18).

This wraps up the substantial part of my discussion of the standard responses as defeater-defeaters. I conclude that the two standard responses to EDAs are ineffective if understood as defeater-defeaters: they do not serve to reinstate the justification of our moral beliefs. In the next chapter, I will discuss the standard responses understood as defeater-deflectors.

3.5 THE DEFEAT ARGUMENT

Let me now conclude my reasoning by summarizing the argument I have developed here. This argument builds on the following assumption:

> ***DEFEATER-DEFEATER ASSUMPTION***: The standard responses try to reinstate the justification of our moral beliefs.

If the standard responses try to reinstate justification of our moral beliefs, then this presumes that this justification was lost at an earlier point in time, where we (the moral believers) were confronted with the information about the evolutionary influence on our moral beliefs. This means that after being confronted with the considerations supporting the generic EDA (call the set of information that comprises these considerations DEBUNKING),[89] our moral beliefs are defeated, and so, at a certain point in time, at t_1, it is the case that we should withhold or disbelieve that our moral beliefs have been reliably formed via our moral faculty.

From here, the following argument unfolds, which I have termed the *Defeat Argument* for easy reference. This argument starts with the following premise: at t_2, the standard responses rely on COMMON-SENSE MORALITY to support the argument for the reliability of our moral faculty. This premise basically just states a crucial assumption, on which the standard responses must rely.

---

[89] In other words, the considerations that we are informed about at a certain point in time (t_1 in the defeater-defeater-case; t_2 in the defeater-deflector-case), and which make it the case that the best explanation available to us concerning our moral faculty does not involve the relevant moral facts, and which make it the case, that we become aware (or should become aware) that we lack independent means for confirming the reliability of our moral beliefs.

At t_1, upon being confronted with DEBUNKING, we should withhold or disbelieve that our moral beliefs have been reliably formed via our moral faculty. This premise states that at a certain point in time (earlier than t_2), we have a(n) (undercutting) defeater for our moral beliefs, which straightforwardly follows from S being in a defeater-defeater-scenario. Furthermore, in the present context, we assume that this undercutting defeater is generated by the incoming information DEBUNKING.

Between t_1 and t_2 nothing else occurs that would make it the case that we should no longer withhold or disbelieve that our moral beliefs have been reliably formed via our moral faculty. This makes explicit that we assume that nothing else occurs, that reinstates the justification of our moral beliefs. We need this assumption for the standard responses to remain relevant.

And now the real action happens: if, at t_1, we should withhold or disbelieve that our moral beliefs have been reliably formed via our moral faculty, and if between t_1 and t_2 nothing else occurs that would make it the case that we should no longer withhold or disbelieve that our moral beliefs have been reliably formed via our moral faculty, then, at t_2, relying on COMMON-SENSE MORALITY to support the reliability of our moral faculty is malignantly epistemically circular. This premise basically applies Bergmann's sufficient condition for malignant epistemic circularity to the present case. It is supported by the above discussion of Bergmann's condition and by the discussion of the cases of Larry and Anne, where the analogous dependence on the truth of defeated beliefs was also intuitively epistemically wrong. It then follows that at t_2, relying on COMMON-SENSE MORALITY to support the reliability of our moral faculty is malignantly epistemically circular.

If a response involves malignant epistemic circularity, then this is sufficient to render that response unsuccessful in reinstating the justification of a belief or of a set of beliefs. This is an assumption that was made during the discussion above. As I have briefly argued, rejecting this assumption seems to conflict with our intuitive application of the notion of undercutting defeat. It then follows that the standard responses are unsuccessful in reinstating the justification for our moral beliefs. This concludes the Defeat Argument.

This result wraps up my discussion of the standard responses as defeater-defeaters. I conclude that the two standard responses to EDAs are ineffective if understood as defeater-defeaters: they do not serve to reinstate the justification of our moral beliefs. In the next section, I will investigate whether the defeater-deflector option is more promising.

4 THE STANDARD RESPONSES AS DEFEATER-DEFLECTORS

4.1 INTRODUCTION

In the last chapter, I have argued that there are good reasons to think that both third-factor accounts and the AMS are unsuccessful as defeater-defeaters. The reason for this is that it is clearly epistemically wrong to continue to assume the truth of your belief in the presence of a defeater for that very belief.

In this chapter, I will now investigate how third-factor responses and the AMS fare if we understand them as defeater-deflectors. If third-factor accounts and the AMS cannot be understood as conditions that defeat a defeater, the hope remaining for the realist is that these responses are successful if understood as providing considerations that prevent the occurrence of potential defeat.

Here is the plan for the chapter: in section (4.2.), I will break down what it means for a reply to be defeater-deflector, and I will determine how we can assess the standard responses as defeater-deflectors. In (4.3), I will revisit the cases of Larry and Anne from the last chapter to determine whether, intuitively, the subjects in those cases ever gain a defeater. In section (4.4), will try to get clear on the epistemological framework that we need to assume, to make the standard responses work as defeater-deflectors. Here I will argue that it seems that the standard responses are in conflict with the NO DEFEATER-condition. In (4.5), I will sketch out a possible motivation for rejecting NO DEFEATER. In (4.6), I will argue that even if the standard responses succeed in protecting our moral beliefs' justification, they are nonetheless deeply unsatisfying, as they recommend a kind of doxastic behaviour that is epistemically vicious. Finally, in (4.7), I will conclude by explicitly stating the argument developed in this chapter.

4.2 THE DEFEATER-DEFLECTOR OPTION

In reply to worries concerning the illegitimacy of third-factor responses, Wielenberg (2014: p. 161 & 2016: p. 506) writes that if third factor-accounts were brought forward as defeater-defeaters (i.e., as reasons reinstating our provisionally lost justification by

countering *prima facie* defeating considerations), they would indeed be suspect. They would be suspect for the simple reason that it is clearly not legitimate to rely on *prima facie* defeated claims to reinstate the justification of those very claims.

But Wielenberg (2016: p. 506) then points to Moon (2016). Moon shows that in contrast to defeater-defeater-cases, it is not necessarily malignantly circular to rely on the truth of your belief that p to deflect an undercutting defeater for your belief that p produced by M. The reason for this is fairly simple. Recall Bergmann's sufficient condition for when epistemic circularity is malignant: if S already has a(n) (undercutting) defeater for her belief that p via M, then it is malignantly circular to continue to rely on her belief that p (or any other belief relevantly dependent on M) in support of the reliability of M.

In defeater-defeater-cases, it is necessarily malignantly circular to rely on the truth of your belief that p, since in those cases it is necessarily true that you have a defeater for your belief that p *qua* this being a defeater-defeater-case. In contrast, in defeater-deflector-cases, it is not necessarily malignantly circular to rely on the truth of your belief that p in support of the belief that M has reliably produced your belief that p. The reason for this is simply that in defeater-deflector-cases, it is not necessarily the case that you already have an undercutting defeater for your belief that p. In other words, it is not necessarily true, that at one point along the timeline described by those cases you should withhold or disbelieve that your belief that p has been reliably formed via M.

The Bergmann/Moon-point is then that benign or malignant epistemic circularity depends on whether the subject in question already should withhold or disbelieve that her belief that p has been reliably formed, and since in defeater-deflector-cases it is possible that the subject should not, it is possible that her dependence on the relevant belief in support of the reliability of M is not malignantly circular. What I mean to say with this is that in some of those cases, continued dependence on a belief, that is the product of M is not malignantly circular, but in some of these cases it is. To support this point, it will be useful to have an example for each kind of case in hand.

Consider the following example, taken from Moon (2016: p. 12):

> ***ZZ COLOUR VISION***. You believe that your colour vision is overall reliable. You walk into a room with objects that have no standard colour (e.g., there are no bananas in the room, but chairs and bowls). You form the belief that the wall is red, the bowls are blue, and so on. Then, a friend, who you know to be an exceptionally reliable testifier, tells you that that the drug ZZ was mixed into your food earlier today. She also tells you that ZZ renders the colour vision of 95% of those who ingest it permanently unreliable.

Is it epistemically unproblematic for you to rely on your beliefs about the colours of the objects in the room to show how the information about ZZ does not undercut the justification for your relevant colour beliefs? Are you entitled to reason in the following way:

> ZZ renders the colour vision of 95% of those who ingest it permanently unreliable. But this wall in front of me is red, and these bowls are blue! So, I must belong to the few people who are immune to ZZ!

In this case, you infer that your colour vision is still reliable from the truth of your colour beliefs. You reason here that ZZ has not rendered your colour vision unreliable, since your colour vision still produces true beliefs, and so you must be immune to ZZ. Your colour beliefs here are meant to function as information that gives you reason to think that the information about ZZ does not undercut the justification of your colour beliefs at the point in time where you receive it. In other words, the truth of your colour beliefs is meant to deflect defeat when you are informed about ZZ. But can your colour beliefs successfully deflect the defeater in this case?

The answer to this question seems to be "No". Your colour beliefs about the objects in the room are not fit to the task of preventing defeat from occurring. The reason why your beliefs about the colours of the objects in the room are not able to prevent the occurrence of defeat is that as soon as your friend gives you the relevant information you instantly have an undercutting defeater for those very beliefs (ibid.). Your colour beliefs do not seem to be capable of preventing the occurrence of defeat. And once again, since those beliefs are defeated as soon as you receive the information about ZZ, relying on those colour beliefs in support of the reliability of

your colour vision is malignantly circular. So, your colour beliefs are inadmissible as defeater-deflectors.[90]

Put a bit more formally, what the case of ZZ COLOUR VISION shows is the following: the possession or reception of a piece of information POTENTIAL D-DEFLECTOR at t_1, will not be sufficient for keeping information D from defeating S's belief at t_2, if before or at t_2, S's justification for believing POTENTIAL D-DEFLECTOR is itself defeated.

For example, assume you have a belief that p via M that is justified at t_0. For a piece of information that you receive or possess at t_1, which is also epistemically dependent on M (in the sense that the relevant beliefs have also been formed via M) to be sufficient for keeping defeat from occurring, it must not be the case that before or at t_2, your justification for believing in the potentially deflecting piece of information is itself defeated. But now assume that the potentially undercutting information you receive at t_2 calls into question the reliability of M generally, i.e. it is now the case that you have an undercutter for all beliefs epistemically dependent on M. This gives you at t_2 an undercutting defeater for belief in POTENTIAL D-DEFLECTOR. Therefore, you must not rely on POTENTIAL D-DEFLECTOR, as relying on defeated beliefs in this way is malignantly circular. You must not already have an undercutter for your belief that p via M, as otherwise relying on beliefs which are epistemically dependent on M is malignantly circular.[91] And as Moon puts it, in the case of ZZ COLOUR VISION, it seems that all beliefs epistemically dependent on your colour vision are undercut in "one fell swoop" (Moon 2016: p. 12).

Now consider a different example, again adapted from Moon (2016: p. 13).

> ***YY IMMUNITY***: You believe that your cognitive faculties are overall reliable. A scientist whom you know to be trustworthy tells you that several highly reliable tests clearly show that you are one of the few who is immune to the effects of YY.

---

[90] If the ZZ COLOUR VISION-case strikes you as similar in some respects to the cases of Larry and Anne, then this is no coincidence: I take this example from Moon (2016: p. 12) who acknowledges that the example is similar to Locke's (2014: p. 231) "Martian"-case, and as I have acknowledged in the last chapter, my presentation of the cases of Larry and Anne draws on Locke's example.

[91] Assuming that the relevant undercutter calls the reliability of M in general into question, as it is the case in the moral belief-case and the colour belief-case.

> You ingest YY. You walk outside and see a red car in the driveway. You form the belief that there is a red car in front of you. You then learn that YY destroys the cognitive reliability of 95% of those who ingest it.

You believe that there is a red car in front of you. You believe that despite ingesting YY, your cognitive abilities are still reliable. The consideration for why you believe that your cognitive abilities are still reliable is itself based on your cognitive abilities (which you use e.g., to register and take in the testimony of the scientist and memorize it, and furthermore, your knowledge about the scientist's trustworthiness is also based on your cognitive abilities). When questioned on why you believe yourself to be reliable, you would offer these considerations, and your reasoning would show that your belief that your cognitive faculties are still reliable exhibits epistemic circularity. In other words, your belief that your cognitive faculties are still reliable is itself based on those very faculties. But in this case, the circularity appears to be benign, as you do not have an undercutting defeater that makes it the case that you should withhold or disbelieve that your cognitive faculties are reliable.

Your justified belief that you are one of the immune 5% seems to be an admissible defeater-deflector for the potential defeater that your cognitive reliability is gone after you have ingested YY. This potential defeater never seems to gain defeating power in the first place in YY IMMUNITY, as it would be odd to hold that believing that you took YY provides you with a reason to doubt the trustworthiness of your cognitive abilities – while you also believe (on good grounds) that you are immune to YY (Moon 2016: pp. 13-14). Given everything that you believe or should believe in this scenario, it is not the case that getting informed that YY destroys the cognitive reliability of 95% of those who ingest it makes it the case that you should withhold or disbelieve that your belief that p was reliably formed via M.

Therefore, your belief that you are one of the immune 5% seems to be an admissible defeater-deflector, fit to protect the justification for your belief that there is a red car in front of you, even though both beliefs have been formed via your cognitive faculties. In line with Bergmann's sufficient condition for malignant epistemic

circularity,[92] this kind of circularity is not problematic, since you lack a defeater for your belief that there is a red car in front of you.

As I have written above in discussing ZZ COLOUR VISION, the question of whether some piece of information is an admissible defeater-deflector importantly depends on whether we remain justified in believing in this potentially deflecting information once we are confronted with the potential defeater. If we remain justified, then this information can be an admissible defeater-deflector. If we don't, it can't. In ZZ COLOUR VISION, you gain an undercutter for beliefs based on your colour vision, and therefore it is malignantly circular to rely on those beliefs. This disqualifies them as defeater-deflectors. In YY IMMUNITY, you do not gain an undercutter for beliefs based on your cognitive faculties, and therefore relying on those beliefs is only epistemically circular in a benign fashion. And so, in YY IMMUNITY, your beliefs about your immunity to YY, although epistemically dependent on your cognitive faculties, seem to be able to deflect the potential defeater.

The important upshot of the discussion of these two examples is that while in potential defeater-deflector-cases, epistemic circularity need not be malignant, there are still such cases where it is malignant. The crucial question we must answer therefore becomes:

> Is it epistemically malignant for the standard responses to rely on COMMON-SENSE MORALITY?

Given Bergmann's sufficient condition for malignant epistemic circularity, this basically means to ask: is it the case that at a certain moment in time[93] the moral believer should withhold or disbelieve that her moral beliefs were formed reliably via her moral faculty? The answer to this question determines directly whether the defeater-deflector-option is any more promising than the defeater-defeater-

---

[92] To remind us, here's what this condition tells us: if S already has a defeater for believing that p is reliably produced by M, then using her belief that p, which is a deliverance of M, as her support for continuing to believe that M is reliable results in malignant epistemic circularity.

[93] That means at or before the point in time at which the moral believer is confronted with the information about the evolutionary explanation of her moral faculty. But since we here are interested in what happens once the moral believer is confronted with the relevant evidence, I will only look at what happens at the point in time, where the moral believer is confronted with the information about the evolutionary explanation for her moral beliefs.

interpretation of the AMS and third-factor accounts was. For if we already have a defeater for our moral beliefs, then our common-sense moral beliefs are just as inadmissible as a defeater-deflectors as are the colour beliefs in ZZ colour vision.

The motivation for thinking of the standard responses as defeater-deflectors lies in the recognition that admitting that an argument like the generic EDA generates an undercutting defeater makes trouble for these responses (for the reasons discussed in the last chapter). In light of the considerations so far, it seems the best way of conceiving the standard responses is as providing us with accounts that are meant to show why an argument like the generic EDA does not give us a defeater for our moral beliefs in the first place. So, it is claimed that the generic EDA never leads to a loss of justification for our moral beliefs.

I have already stated that with Wielenberg, we have at least one third-factor theorist who explicitly goes for this option. This reading is also in line with how Clarke-Doane presents the AMS. As a response to the generic EDA, the AMS basically tells us that the presence of a certain explanation of our moral faculty is insufficient for establishing that we should withhold or disbelieve that our moral beliefs were formed reliably without giving us some reason to believe that our moral beliefs are either not safe or not sensitive. Given MODAL SECURITY, if your basic moral beliefs are true and if there is a modally-robust explanation for why you hold the moral beliefs you do,[94] then the evolutionary explanation of your moral faculty does not take away your justification for holding your moral beliefs. The presence of the evolutionary explanation does not make it the case that you should withhold or disbelieve that your moral faculty is reliable without giving you reason to think that the beliefs produced by this faculty are not safe or not sensitive. If your basic moral beliefs are true, and if you could not easily have had different beliefs, then the presence of the evolutionary explanation does not and cannot show you to be unreliable. This is how the AMS is meant to counter the explanatory challenge seemingly arising from EDAs: by showing it to be illusory.

[94] That means to say, an explanation that shows that you could not easily have believed differently.

So, the standard responses seem to argue that arguments like the generic EDA at no point generate an undercutting defeater for our moral beliefs. The question that we must discuss therefore is whether the considerations offered by the standard responses are sufficient to prevent the occurrence of defeat.

4.3 THE CASES OF LARRY AND ANNE REVISITED

In the last chapter, I have presented the cases of Larry and Anne. These are examples of cases where, intuitively, it would be epistemically wrong for the subjects to hold onto their beliefs in the face of certain incoming information. Intuitively, neither Larry's belief in God nor Anne's belief in spirit animals is justified after they receive information about what best explains their way of arriving at those beliefs.

I have used these examples to show that if a subject's relevant beliefs are defeated, then it would be epistemically wrong for these subjects to continue to hold and rely on those beliefs. The question now under discussion is different. It is not the question of whether one can rely on the truth of a defeated belief to support the claim that this belief was formed reliably. It is the question of whether moral believers who receive the relevant information about the evolutionary influence on our moral belief-forming faculty ever gain an undercutting defeater for their moral beliefs. You might therefore object to my continued use of the examples of Larry and Anne.

But for the point I want to make here, the examples of Larry and Anne do seem instructive and fitting: they are cases, which are (or so I have argued) exactly analogous with respect to all the relevant features to the cases of the respective moral believers.

We now want to determine, whether the relevant moral believers gain a defeater for their moral beliefs upon receiving a certain kind of information, or whether the considerations involved in the AMS and third-factor accounts are sufficient for preventing the occurrence of defeat. And for the question of whether the relevant moral believers have an undercutting defeater for their moral beliefs at that point in time, it does seem relevant to ask whether we intuitively judge that subjects in exactly analogous cases have an undercutting defeater. If the subjects' beliefs in the analogous cases are defeated, it seems, so are the relevant moral beliefs. On the other hand, if it is the case that the relevant moral believers have an admissible defeater-deflector available to them, then, given that these cases are exactly analogous to the cases of

Larry and Anne, we should also expect that Larry and Anne have an admissible defeater-deflector available to them.

In the last chapter, we were concerned with the question: given that your belief is undercut, why is it wrong to rely on an undercut belief in support of the reliability of your method for arriving at those beliefs? Now we are concerned with the question: in the relevant moral belief-cases, are our common-sense moral beliefs undercut? And for answering this second question, our verdict about cases which are exactly analogous to the relevant moral belief-cases is relevant.

At t_0, Larry is justified in believing that God exists via his intellectual seeming, and Anne is justified to believe that Jordan's spirit animal is an owl. Given everything else they believe, is it the case, that at t_2 (i.e., the point in time at which they are confronted with the information about what best explains their respective intellectual seemings and become aware that they lack the means for independently confirming the reliability of their respective belief-forming faculties) they should withhold or disbelieve that their relevant beliefs are reliably formed via their intellectual seemings?

Here I think we must note that, intuitively, this seems to be the case. It seems Larry and Anne should each withhold or disbelieve that their relevant beliefs were formed reliably via their intellectual seemings. Given the cases of Larry and Anne, which were presented in the last section, we should be immediately suspicious of the claim that the considerations offered by the standard responses could be sufficient for blocking the occurrence of defeat.

Part of my argument in the last chapter was that we can find cases, which are exactly analogous to the cases of the relevant moral believers and where the beliefs of the subjects in those cases are intuitively defeated – even though these subjects can reason that if their beliefs are true, then they are modally secure or not subject to a problematic kind of epistemic luck. In the two cases discussed earlier, considerations which are exactly analogous to the considerations involved in the generic EDA seemingly were intuitively sufficient to defeat the justification of Larry to believe in God and of Anne to believe in spirit animals.

To work as defeater-deflectors, the standard responses must hold that the believers in the relevant moral belief-cases at t_2 do not have an undercutting defeater for their

moral beliefs due to evolution, because the considerations offered by these responses stop the occurrence of defeat. To deflect the defeater, both responses need to rely on the first-order moral claims contained in COMMON-SENSE MORALITY. But belief in those first-order moral claims is only legitimate in support of the reliability of the believer's moral faculty, if it is not the case that the beliefs epistemically dependent on the moral faculty are undercut (cf. Crow 2016: pp. 390-391).

The cases of Larry and Anne seem to provide as with examples of cases, which are (in all relevant respects) exactly alike the cases of the relevant moral believers, and where considerations in terms of modal security or a third-factor explanation intuitively do not suffice to prevent the occurrence of defeat. They are not sufficient for preventing the occurrence of defeat, since

(i). the considerations offered by these responses are dependent on the truth of relevant first-order claims (COMMON-SENSE MORALITY in the moral belief-cases; the claim that God exists in Larry's case; the claim that spirit animals exist in Anne's case),

(ii). belief in these first-order claims is epistemically dependent on the very faculty, whose reliability is under debate,

(iii). and, intuitively, in the exactly analogous cases of Larry and Anne, the information received by the subjects serves to undercut the justification for all beliefs epistemically dependent on the relevant faculty.

Given (iii) and given the exact analogy between the cases of Larry and Anne and the relevant moral belief-cases (in all epistemically relevant respects), it seems that in the moral belief-cases too, all beliefs epistemically dependent on the moral faculty are undercut in one fell swoop.[95]

[95] This result is very similar to a point made by Crow (2016: pp. 390-391). As Crow points out with respect to third-factor explanations, one's justification for believing e.g. that the goodness of survival is partially metaphysically dependent on the facts about survival or that facts about cognitive faculties partially ground facts about rights depends on one's justification for believing particular first-order moral claims. So, the justification for believing the relevant explanatory claim in the third-factor accounts depends epistemically on first-order moral claims. And assuming the empirical hypotheses involved in EDAs are correct, our beliefs in those claims depend causally on evolutionary pressures.

This leaves the standard responses with two choices: either to argue that the cases of the relevant moral believers are not analogous to the cases of Larry and Anne, or to argue that the beliefs of Larry or Anne are not actually defeated by the incoming information about the origin of those beliefs since Larry or Anne can deflect the defeater.

As I have argued in the last chapter, it seems we have good reason to think that the cases of Larry and Anne are exactly analogous in all relevant respects to the cases of the moral believers. Absent any convincing argument for the claim that there is a relevant disanalogy, it appears plausible that this is indeed true.

4.4 The Epistemology of the Standard Responses as Defeater-Deflectors

This leaves open only the second option. That is to argue that Larry or Anne have after all access to admissible defeater-deflectors, and so, at the relevant point in time t_2, it is not the case that Larry or Anne already have or gain an undercutter for their respective beliefs.

Holding that Larry's and Anne's beliefs are not undercut by being informed about the drug and the tumour seems to be rather counter-intuitive. Indeed, given all I have said so far, it seems that the cases of Larry and Anne are clear counter-examples to the epistemological stories undergirding the AMS and third-factor accounts respectively. But at this point, Clarke-Doane and the third-factor theorists would perhaps simply insist that the beliefs of Larry and Anne respectively are never defeated due to the presence of admissible defeater-deflectors. Since I am interested in making progress, I won't simply stand my ground on this issue. Before I give my further argument for the claim that even if Larry's and Anne's beliefs are justified, their doxastic behaviour[96] is nonetheless epistemically criticisable, I think we should try to answer the following question:

---

[96] What is "doxastic behaviour"? I assume that we do not usually have voluntary control over our beliefs (or over our doxastic attitudes more generally). In other words, we cannot usually choose what doxastic attitude to adopt towards a proposition. Nonetheless, as Peels (2017: p. 2898) writes, we have direct or indirect control over "belief-influencing" activities, activities which taken together comprise what I call here "doxastic behaviour". These are e.g. "evidence gathering, working on our epistemic virtues and vices, and improving the

> What could be the epistemological motivation for claiming that Larry's and Anne's beliefs are not undercut?

That means to ask: what kind of epistemological story do you have to suppose to arrive at the result that Larry's or Anne's beliefs do not lose their justification at t_2?

Both standard responses seem to rest on an externalist conception of epistemic justification, according to which justified belief is a standing "one has in virtue of the (de facto) reliability of the processes one employed in arriving at truth" (Goldberg 2014: p. 280).[97] The AMS also explicitly comes with an account of undercutting defeat that seems to fit an externalist conception of justification: information E can only undercut the justification for S's belief that p, if it gives S reason to think that her belief that p is either not safe or not sensitive.[98] Given this account of undercutting defeat, Clarke-Doane might now stand his ground in the face of my reasoning in this section so far, and state that Larry's belief is not undercut at t_2 since Larry himself sees that the information about the drugs does not show on its own that his beliefs were either not safe or not sensitive.

Although third-factor accounts do not make this explicit, it seems that they too would have to endorse a similar account of undercutting defeat in order to motivate the claim that Anne's belief in spirit animals is not defeated (along the lines of: information E can only undercut the justification for S's belief that p, if it gives S reason to think that her relevant belief was subject to epistemic luck in a problematic sort of way).

So, given that the standard responses need some epistemological motivation for their claims that the beliefs of Larry and Anne respectively are not undercut, it seems they are committed to one of these externalist accounts of undercutting defeat (and in the

---

functioning of our doxastic mechanisms—briefly, our evidence bases and our belief-forming habits" (ibid.). Since subjects' in many cases have sufficient control over their doxastic behaviour understood in this sense, it might be appropriate to hold them (intellectually) responsible for these belief-influencing activities.

[97] For a few paragraphs on the distinction between internalism and externalism about epistemic justification, please see appendix A.

[98] Internalists about epistemic justification could certainly agree that your belief can become unjustified if you receive information that indicates that your belief is not safe or not sensitive. But internalists typically would not restrict themselves to stating that only information like this can make your belief become unjustified via undercutting it.

case of the AMS, this commitment is even an explicit and integral part of Clarke-Doane's argument). Call these kinds of views, which both roughly hold that some information E can only undercut the justification for S's belief that p, if it gives S reason to think that her relevant belief is not *modally* reliable, "strongly externalist views of undercutting defeat". According to these views, information E can only undercut S's belief that p via showing that the belief has been subject to a problematic kind of epistemic luck.

If you commit yourself to one of these accounts of undercutting defeat, you will hold that Larry's or Anne's justification for holding their respective beliefs remains intact even after receiving the information about what best explains their intellectual seemings (and after becoming aware that they lack independent means of confirmation). Part of the argument of the last section was that Larry and Anne both could and should take a negative perspective toward whether their relevant beliefs have been produced in a reliable fashion. In other words, from their perspective, it seems perfectly rational to take their own beliefs to be undercut by the incoming information. But the important upshot of committing yourself to a strongly externalist kind of undercutting defeat is that the fact that from your own perspective, it is perfectly rational to take your belief to be undercut does not make it the case that your belief is undercut. The fact that you rationally believe that your belief that p was not reliably formed via M does not suffice to make it the case that you should withhold or disbelieve that your belief that p was reliably formed via M.

Let us define some terminology, to get clearer into view what's now at issue. Call a "rationality undercutter" any information, which upon its reception by a believer makes it the case, that the believer is epistemically irrational if she does not take her belief to be undercut.[99] Importantly, a rationality undercutter may not give the believer

---

[99] The notion of a "rationality undercutter" developed here is built to accommodate the possibility that epistemic rationality and justified belief might come apart. We need to accommodate this possibility to make sense of the position I have termed strong externalism about undercutting defeat. Of course, you might just deny that rationality and justified belief can come apart, because you think that "rational belief" and "justified belief" are synonyms or close to being synonyms. I have great sympathies with that response. But I abstain from making this reply, because I am interested in further exploring the epistemological story that supports the standard responses. Here, I try to give these responses their best shot. Ultimately, I think that this makes the line of reasoning developed here more damaging to the standard responses, as I hope to show that even buying into a highly controversial

any reason to think that her belief is not *modally* reliable (which is demonstrated by the cases of Larry and Anne, where the subjects believe that their relevant beliefs were formed *modally* reliably, if their beliefs are true and where we assume that their beliefs are true). Therefore, accepting one of the above externalist views on undercutting defeat has the consequence that rationality undercutters are no undercutters at all.

Undercutting defeat, according to these views, is not so much about what epistemic perspective it is rational for a believer to adopt towards her own beliefs from her own limited, first-personal point of view as a believer. Rather, it is exclusively about whether the relevant information gives the believer any reason to think that her belief was subject to a problematic sort of epistemic luck.

What's characteristic about strong externalism and important for our purposes is that this account holds that the possession of a rationality undercutter is not sufficient for generating an undercutting defeater for S's belief that p. This is important, as you might have a rationality undercutter for your belief that p even though you see that if your belief is true you could not have failed to arrive at the truth or that your believing truly is not a matter of problematic epistemic luck (this is supported by the cases of Larry and Anne). Assuming strong externalism about undercutting defeat then allows you to say that the only way for your belief to be undercut is by receiving information that indicates that your belief is not *modally* reliable. But that's not something that a rationality undercutter necessarily indicates (which is again intuitively attested by the cases of Larry and Anne).

Strong externalism about epistemic defeat is therefore the view that is open to the possibility that S's belief is not undercut (due to the fact that S's circumstances are such that they guarantee that her belief is modally secure or not true as a matter of problematic epistemic luck, *if* her belief is true), even though judging from the perspective of the relevant believer, we would, intuitively, say that that believer should rationally take her belief to be undercut.[100]

---

epistemological account does not achieve the goal of making these responses work satisfyingly.

[100] The view that I call "strong externalism about epistemic defeat" is therefore, it seems, at odds with the intuitive thought that underlies the NO DEFEATER-condition, i.e. the thought epistemic justification has a perspectival dimension. More on this below.

The benchmark that undercutting defeaters therefore must meet (according to the strong externalist views) is that they must give the relevant believer sufficient reason to think that her belief is *modally* unreliably formed via M even while assuming that her belief is true.

Importantly, this benchmark is not met in either the cases of Larry or Anne, where, *ex hypothesi*, the two subjects (justifiably) believe that their relevant beliefs are *modally* reliably formed if they are true. And in those cases, it also clear that the information they are receiving (about the drugs and the tumour respectively) do not give them any direct reason to think that their relevant beliefs are false. It therefore seems that (assuming strong externalism about epistemic defeat) Larry's and Anne's respective beliefs might well not be defeated at t_2, when they receive the relevant information (about the drugs and the tumour respectively).

It is noteworthy, that many externalists about epistemic justification are not strong externalists about undercutting defeat. Many externalists accept the proviso that purely internal evidence can function as a defeater for a belief's justification even if this belief was originally justified in virtue of external features alone (e.g., via being the product of a *de facto* reliable mechanism) (Baker-Hytch 2017: p. 6; cf. Bergmann 2006: Ch. 6; Goldberg 2014; Sudduth 2008: section 3a). These externalists about epistemic justification agree with internalist views at least on the claim that that for S's belief to be undercut, it suffices if S, given her own epistemic perspective on the situation, takes or should rationally take her belief to be undercut.

Why would externalists about epistemic justification build this proviso into their accounts? A reason that strikes me as a good one is that this allows you to give a straightforward account of why subjects like Larry and Anne, who rationally should take their belief to be undercut, lose their justification for holding their respective beliefs, even though they have not gained a reason to think that their beliefs are *modally* unreliable.[101] On the other hand, commitment to a strongly externalist account of

[101] My remarks here are in line with Goldberg, when he writes that "not all de facto reliable processes are such that the subject is entitled to rely on them" (Goldberg 2014: p. 292). To demonstrate this point, Goldberg uses a version of BonJour's (1980) famous example of a *de facto* reliable clairvoyant. My examples of Larry and Anne are like this clairvoyance-case insofar, as they are all trying to establish the point that a subject can base her beliefs on a *de facto* reliable process, and yet she would be, intuitively, epistemically irrational to believe or

undercutting defeat of the above sort has the implication that subjects (like Larry or Anne) can retain justification for their beliefs – when, intuitively, (i) the justification for their beliefs is undercut, and (ii) the subjects themselves could and should think of their own beliefs as being undercut.

But assume that either the AMS or third-factor accounts (or both) present us with successful defeater-deflectors. This also means that the beliefs of subjects like Larry and/or Anne remain justified, even though both Larry and Anne have a rationality undercutter for their beliefs. According to strong externalism about undercutting defeat, this rationality undercutter is "misleading". That means it does not give the relevant subjects reason to believe that their beliefs are not *modally* reliable, and so does not undercut the justification for holding those beliefs.

At this point, we need to tread carefully, and make some subtle, but important points to get clear on what is at issue here. I have stated that subjects like Larry and Anne have a rationality undercutter for their relevant beliefs. In other words, they are epistemically irrational if they do not take their beliefs to be undercut. Given that Larry's and Anne's cases are exactly analogous to the case of the relevant moral believers, it seems that those moral believers also at least have a rationality undercutter for their beliefs.

Here's how I would say that a rationality undercutter gives rise to an undercutting defeater. A rationality undercutter makes it the case that given her doxastic/normative set, S should withhold or disbelieve that p*. And this undercuts S's justification for holding her belief that p. To me this seems to be an intuitive account of how epistemic rationality and epistemic justification are connected. This account also seems perfectly in line with the epistemologists who support versions of the NO DEFEATER-condition.[102] But in some way, the kind of strong externalism about undercutting defeat, to which the standard responses are committed, seems to be at odds with this intuitive and popular account.

---

keep believing on the basis of this information given everything else that she believes or should believe.

[102] Cf. Bergmann (2006: Ch. 6).

There are three possibilities here on how strong externalism conflicts with this picture or might seem to conflict with this picture. The first possibility does not consist in a very deep disagreement: you could assert that this general picture is right, but state that it does not apply to the cases discussed. It is therefore not at all a disagreement with the above picture as such, but just with the application of the picture to certain cases. That means you could deny that subjects like Larry and Anne have a rationality undercutter.

But that seems to commit you to saying that there is nothing epistemically wrong with subjects like Larry and Anne. That means not only is it the case that their justification remains intact, they are not even in the least bit epistemically criticisable for displaying their relevant doxastic behaviour. I should make explicit that my own explanation for why Larry and Anne's respective doxastic behaviour would be epistemically criticisable below will depend on the assumption that they do have a rationality undercutter for their relevant beliefs. In this argument, the explanation of why subjects like Larry and Anne are behaving in an epistemically wrong kind of way assumes the presence of a rationality undercutter. And so, in my explanation, the epistemic wrongness of their conduct in retaining their relevant beliefs is dependent on their possession of a rationality undercutter. So, by denying that Larry and Anne have a rationality undercutter, you can undermine my argument.

That being said, it seems rather incredible to state that subjects like Larry and Anne do not have a rationality undercutter and are therefore not rightly epistemically criticisable if they retain their original beliefs. It just seems very intuitive to think that Larry and Anne should rationally think of their beliefs as being undercut by the information about the drugs and the tumour respectively. I therefore contend that every account of epistemic rationality that supports the result that subjects like Larry and Anne should not rationally take their beliefs to be defeated owes us a very good argument. (What I mean to say is that such accounts owe us a very good argument to the effect that subjects like Larry and Anne should not rationally take their beliefs to be defeated.) Absent such a very good argument, we should therefore reject the claim that subjects like Larry and Anne should not rationally take the relevant beliefs of theirs to be defeated.

The second option is that you can concede that the relevant subjects have a rationality undercutter, but state that the possession of a rationality undercutter does not make it the case that, given your doxastic/normative set, you should withhold or disbelieve p*. In other words, the presence of a rationality undercutter does not make it the case that you have an undercutting defeater, since a rationality undercutter does not make it the case that you adopt or should adopt a negative perspective towards your belief.

The problem I see here is that if the presence of a rationality undercutter is not sufficient to make it the case that, given your doxastic/normative set, you should withhold or disbelieve that p*, then I am not sure what is. The second option concedes that subjects like Larry and Anne have a rationality undercutter. That means that Larry and Anne are epistemically irrational if they do not take their beliefs to be undercut. It seems eminently plausible to say that this makes it the case that they should withhold or disbelieve that their relevant beliefs were formed reliably. If it is the case that from your own perspective, it would be irrational to hold onto a belief in the presence of the information you are now receiving, then, it seems, given everything you believe or should believe, you should withhold or disbelieve that your belief was formed reliably.

For example, intuitively, in ZZ COLOUR VISION, it is the case that your belief that the wall is red gains a rationality undercutter. In that case you are told that you have ingested ZZ before you came into the room – and this makes it the case that you would be epistemically irrational, if you did not adopt the view, that the justification for your belief that the wall is red is undercut. By contrast, in YY IMMUNITY, we can partly account for why you shouldn't withhold or disbelieve that your cognitive faculties are reliable via pointing out that you never seem to gain a rationality undercutter in this scenario. What this comparison between ZZ COLOUR VISION and YY IMMUNITY suggests is that the presence of a rationality undercutter for your belief that p seems to be sufficient for making it the case that, given your doxastic/normative set, you should withhold or disbelieve that p*. Again, absent a very good argument against this claim, I think we should accept it.

Third, you can concede that the subjects have a rationality undercutter and concede that the subjects now, given their doxastic/normative sets, should withhold or disbelieve p*. But you could deny that the fact that you should withhold or disbelieve

that p* (given your doxastic/normative set) makes it the case that you lose your justification for believing that p. That means denying the NO DEFEATER-condition on epistemic justification. Given everything we have said so far, this is the only option left.

The surprising sub-conclusion we have now arrived at is that the standard responses are committed to the rejection of the NO DEFEATER-condition on epistemic justification. We have arrived at this sub-conclusion by noticing that the standard responses are best seen as defeater-deflectors. Their success as defeater-deflectors is dependent on the claim that before or at t_2, our moral beliefs are not undercut. But in the exactly analogous cases of Larry and Anne, the subjects in those cases have an undercutting defeater for their respective beliefs at t_2 – or at least, that's what it looks like.

Now, the standard responses need to state that Larry or Anne's beliefs at t_2 are not undercut if they want to avoid stating that the beliefs of the moral believers in the analogous cases are undercut. Here they meet the complication that it seems that they must concede that Larry and Anne have a rationality undercutter, and that a rationality undercutter makes it the case that the subjects should withhold or disbelieve that their relevant beliefs were reliably formed. In response to this, they must hold that it is not necessary for S's belief that p to be justified that S, given her doxastic/normative set, should not withhold or disbelieve that p*.

At this point, I suspect that many epistemologists would stop the argument and simply state that this suffices for us to reject the standard responses.[103] In section (1.4.1), I have stated that I think it is safe to say that the NO DEFEATER-condition seems to be widely endorsed. If the standard responses need to reject this condition (as it seems they must) on account of being committed to a kind of strong externalism about undercutting defeat, we should ask whether there is any independent support in the literature for this kind of view on what gives rise to a loss of justification after

[103] If we assume that NO DEFEATER (or a relevantly similar version of it) spells out a necessary condition for justified belief in a proposition, then we are able to stop the argument at this point. The argument against the standard responses as defeater-deflectors, conditional on the truth of the NO DEFEATER-condition (or a relevantly similar version of it) would then simply be that the cases of Larry and Anne show that the considerations offered by the standard responses are insufficient to prevent the occurrence of defeat.

receiving information that changes what you believe or should believe.[104] Why would you reject the NO DEFEATER-condition on epistemic justification? I will investigate this question in the next section.

4.5 EXTERNALIST SCEPTICISM ABOUT DEFEAT

Maria Lasonen-Aarnio (2010) has argued that epistemic externalists have reason to be sceptical of "no defeater"-conditions on justification and knowledge. She argues that a subject's knowledge and justification can be retained even in cases where this subject is confronted with evidence that intuitively makes it the case that the subject should revise her relevant beliefs, and where the subject would be intuitively epistemically criticisable for failing to do so.[105]

Why should we think that Lasonen-Aarnio's arguments can provide support for the crucial point that S's possession of a relevant rationality undercutter is not sufficient for taking away the justification of S's belief that p? Assume the above strong externalism about epistemic justification. Given that one accepts this view on justification, one is now confronted with the problem that there seem to be cases where a subject's belief is epistemically negatively affected by incoming information, but where the information that this subject receives does not imply that the subject's belief is not *modally* reliably formed. The cases of Larry and Anne are good examples. How should the externalist react to these cases?

One option already sketched out is to accept the internalist proviso above. But that option is not open to the externalists we have in mind here (for taking this option means admitting that the justification for the relevant beliefs is lost in cases like Larry's and Anne's). Lasonen-Aarnio wants to make room for the option that subjects in such cases can preserve the justification of their beliefs. And this is just what the

---

[104] Due to the apparent conflict between the popular NO DEFEATER-condition on epistemic justification and strong externalism about epistemic defeat, another fitting label for this view (which externalist supporters of the NO DEFEATER-condition might prefer) is perhaps "externalistically motivated scepticism about undercutting defeat".

[105] The way in which I relate Lasonen-Aarnio's arguments to the debate between the evolutionary debunker and the moral non-naturalist is similar to the way in which Law (2016) relates Lasonen-Aarnio's arguments to the debate between the religious debunker and the defender of religious belief.

philosophers supporting the AMS or third-factor accounts should be looking for, if what I have said so far is right.

Therefore, I suggest here that the philosophers presenting the standard responses could perhaps look towards Lasonen-Aarnio's arguments for support for the point that the presence of a rationality undercutter does not suffice for the loss of justification for the relevant beliefs. Lasonen-Aarnio's reasoning potentially provides the standard responses with an epistemological story on why subjects like Larry or Anne (and subjects in analogous cases, like the relevant moral believers) at t_2 do not lose their justification for their relevant beliefs. As we have seen, this point is highly important: the defeater-deflector option is viable only on the assumption that the moral beliefs which are epistemically dependent on the believers' moral faculty do not lose their justification at t_2 due to the incoming information.

Here's a very rough sketch of Lasonen-Aarnio's reasoning. Suppose Larry's belief that God exists is *modally* reliably formed, i.e. let's say it is true, and it is safe and sensitive if it is true, and so it is both safe and sensitive. Given how I have set up the case, there is no reason to think that the seemingly undercutting information he receives indicates that his belief is unsafe or insensitive, if his belief is true. Assume (just for the sake of the example) that knowledge is just safe and sensitive belief. Then, from the perspective of the (strong) externalist, it actually becomes puzzling why possessing this seemingly undercutting information should make a difference with regards to the justification of Larry's belief. If Larry's safe and sensitive belief constitutes knowledge in cases where he does not possess this information, and if this information does not actually connect with the features of the situation that are alone epistemically salient from the externalist's perspective, how can the presence of this information rob Larry of his knowledge (via taking away his justification)? It seems therefore, that from the (strong) externalist's point of view, there is some reason to think that the presence of this information does not rob Larry of his justification for believing that God exists.

If this is accurate, then the externalist, it seems, has reason to believe that subjects like Larry (or Anne, depending on the details of the externalist view you want to go for) can retain knowledge in cases where this seemingly defeating information is present. And so, subjects like Larry and Anne might be able to retain knowledge (and in turn, justified belief) in cases where they have a rationality undercutter, and where they

therefore should withhold or disbelieve that their way of arriving at their belief is reliable.

So, assume a package of strongly externalists views on justification and defeat, according to which subjects are justified or unjustified depending on whether their beliefs are based on *de facto modally* reliable faculties, and where information only takes away your justification if it indicates that your beliefs are not *modally* reliably formed. Assuming this strongly externalist package of views, Lasonen-Aarnio now provides you with a possible solution on how to deal with otherwise rather inconvenient cases: cases where a subject's belief is *de facto modally* reliable, and where the subject receives some piece of information, that (i) does not indicate that the subject's belief is not *modally* reliably formed, but (ii) where the subject would be intuitively irrational for not taking her belief to be undercut. These cases are inconvenient precisely because the information does not indicate that the belief is not *modally* reliably formed, but, intuitively, it seems that this information can nonetheless take away the justification of the subject's belief. Lasonen-Aarnio provides a principled account for why epistemic justification is not lost in those cases, *if* the above strongly externalist perspective is true.

So, assume that the standard responses co-opt Lasonen-Aarnio's account to explain why the possession of a rationality undercutter for your belief that p is not sufficient for taking away your justification for believing that p. This would allow them to give an account for why it is not the case that Larry's or Anne's justification for holding their respective beliefs is taken away – and so neither is the justification of the moral beliefs in the analogous case, where the relevant subjects are confronted with the considerations powering the generic EDA.

4.6 WHY THE DEFEATER-DEFLECTOR OPTION IS UNSATISFYING

The trouble now is of course that (as Lasonen-Aarnio quite clearly appreciates) it would be nonetheless implausible to declare that the presence of a rationality undercutter is of no epistemic consequence. Lasonen-Aarnio thinks that although a subject's *belief* in such a case need not become *unjustified*, a *subject* who retains her belief in the face of evidence that intuitively makes it epistemically irrational for her to continue to hold this belief is epistemically criticisable. This subject is epistemically

criticisable on account of being *unreasonable.*[106] We are now in a position to see that the explanation from the last section for why it is epistemically wrong for subjects like Larry and Anne to hold onto their beliefs still works – even if we assume that a believer's *unreasonableness* does not preclude a belief's *justification*.

Here's the bare bones version of the argument to follow: if you divorce epistemic rationality from epistemic justification, then that might allow you to hold that subjects like Larry and Anne remain epistemically justified in holding onto their beliefs, despite being (in a clear sense) epistemically irrational for continuing to believe as they do in their respective circumstances. But being epistemically irrational is nonetheless epistemically bad! So, even in a kind of best case outcome for the standard responses, these responses seem capable of deflecting defeat only at the prize of endorsing epistemic irrationality.

In what follows, I will put meat onto the bare bones structure of this argument via presenting a virtue epistemological account of epistemic rationality suggested by Lasonen-Aarnio herself. I do not think that this is the only way of fleshing out what is epistemically wrong with subjects like Larry and Anne, even assuming the justification of their beliefs remains intact. Thus, holding that something is epistemically wrong with subjects like Larry and Anne does not commit you to the specific story to follow.

Nevertheless, I think the account suggested by Lasonen-Aarnio is an attractive one for our present purposes, in that it nicely highlights what I take to be the basic reason

---

[106] Lasonen-Aarnio's general point seems to be the following: she distinguishes between two dimensions of normative (epistemic) evaluation, *success* and *competence*. Let's assume that "knowledge" is the relevant success in the epistemic domain. Many paradigm examples of knowledge involve cases of competent success (e.g., people who know because they exhibit epistemic virtue). But the two dimensions of normative evaluation do not always over-lap: there are cases of incompetent success, and competent failure (e.g., cases where people ignore/take into account misleading evidence, and thereby succeed/fail to know). Amia Srinivasan (2015) puts the same general point regarding normative evaluations this way: there is no guaranteed alignment between the facts about what we are permitted to do or what we are obligated to do on the one hand, and the facts about whether we are blameworthy for what we do on the other – even in cases where non-culpable normative ignorance and incapacity are not involved. All norms can be violated through bad luck and conformed to by good luck. That means there are cases of blameworthy right-doing and blameless wrong-doing. What Lasonen-Aarnio calls "unreasonable knowledge" is a prime instance of incompetent success or blameworthy right-doing.

for why the standard responses are deeply unsatisfying, even assuming they are successful in protecting the justification of our moral beliefs. This reason is that the standard responses license a kind of doxastic behaviour that allows subjects to insulate their beliefs from some piece of information, but where this doxastic behaviour is at odds with how a subject with an intellectually impeccable character would proceed in the same situation.

If we divorce epistemic rationality or reasonableness from justification, why should we continue to think that reasonableness is epistemically valuable? The reasonableness of a believer seems to be epistemically valuable because of its tight connection to the goal of inquiry: knowledge. Consider the following example (Lasonen-Aarnio 2010: pp. 14-15; cf. Law 2016: p. 6). Say Joelle adopts the following rule or policy of belief-formation: Believe that p when you see that p, even in the presence of good evidence for thinking that your senses are not to be trusted. If we assume that seeing that p entails p,[107] this method, when correctly applied, cannot produce a false belief. But nonetheless, Joelle should not adopt this rule, because it results in a bad belief-forming disposition, since a "subject who adopts this method is also disposed to believe p when she merely seems to see that p in the presence of evidence for thinking that her senses are not to be trusted" (Lasonen-Aarnio 2010: pp. 14-15).[108] So, although the above rule cannot result in belief in a falsehood in cases where it is correctly applied, if Joelle were to adopt it, she would also be disposed to stick with her beliefs in cases where her evidence that her senses are untrustworthy is not misleading. So, adopting this rule would result in a disposition that overall is not knowledge-conducive. Assume that adopting this rule gives you knowledge in some cases. If it gives you knowledge in those, it also means that your belief is justified, or at least justified to the degree necessary for knowledge – which is really all the justification we can ask for. Nonetheless, being disposed to ignore good evidence (or seemingly good evidence) for thinking that your senses are not to be trusted (like

[107] For example, because we assume epistemological disjunctivism about veridical perception to be true (cf. Pritchard 2012b).

[108] The important point about this example is that although the above rule, when correctly applied, cannot result in belief in a falsehood, ordinary epistemic agents like us are quite incapable of adopting this rule in our doxastic behaviour without then also being disposed to apply it in cases where it only seems to us that we see that p.

Joelle) is unreasonable, due to the consequences for implementing this policy in one's doxastic behaviour.

As Baker-Hytch points out, this also accounts for "our reluctance as onlookers to *attribute* knowledge in such cases,… [which] is explained by our wish not to reward the subjects in such cases for their employment of an unreasonable belief-formation policy—a policy that *in general* does not yield knowledge" (2014: p. 176). Since plausibly, one major function of ascriptions of positive epistemic statuses like knowledge and justification is to mark other people as trustworthy potential sources of information (cf. Dogramaci 2012 & 2015), we are reluctant to ascribe you the statuses of knowledge and justification if an overall non-knowledge-conducive disposition is involved in the production of your belief or in your continued acceptance of it.

Subjects like Larry and Anne seem to have a rationality undercutter for their relevant beliefs. I want to preclude a potential source of confusion here: the unreasonableness of S does not explain why S has a rationality undercutter.[109] It explains why, in the presence of a rationality undercutter, it would be epistemically wrong to continue to hold the belief that is subject to a rationality undercutter. This of course, assumes that S does have a rationality defeater to begin with. So, to be perfectly explicit about this: I basically just assume that Larry and Anne indeed do have a rationality undercutter for their relevant beliefs. I am not sure that there is a deep explanation for why it would be irrational for Larry and Anne to not adopt the view that their relevant beliefs are undercut, and in any case, I do not pretend to have a deep explanation for this.[110] As stated before, it just seems intuitively obvious. You can reject my argument via rejecting the intuition that backs up this assumption. This assumption is therefore an important limitation of my argument. However, rejecting this intuition and the assumption it backs up seems to come at the cost of having to declare that subjects

---

[109] Although S's unreasonableness does explain why it is epistemically bad for subjects like Larry and Anne to retain their beliefs in the presence of a rationality undercutter for those very beliefs – even though this rationality undercutter does not indicate that their beliefs are *modally* unreliable. In other words, the explanation in terms of unreasonableness accounts (or at least partly accounts) for why it is epistemically bad to hold onto a rationally undercut belief.

[110] A potentially promising explanation is perhaps the one offered by Baker-Hytch and Benton (2015: pp. 56-58). Cf. FN 111.

like Larry and Anne are not epistemically criticisable at all for retaining their beliefs in the respective scenarios. This move seems to force you to state that the doxastic behaviour of subjects like Larry and Anne is without blemish. This seems like a high cost to incur.[111]

With that out of the way, I assume that subjects like Larry and Anne have a rationality undercutter. Plausibly, the presence of a rationality undercutter for your belief that p is sufficient for making it the case that, given your doxastic/normative set, you should withhold or disbelieve p*. Currently, we are discussing the question of why (i) even assuming that the fact that you should withhold or disbelieve that p* does not take away your justification for believing that p, (ii) you are nonetheless epistemically criticisable for believing that p (given that you should withhold or disbelieve that p*).

Subjects like Larry and Anne are epistemically irrational in sticking to their beliefs in the presence of evidence that makes it the case that they should withhold or disbelieve that their relevant beliefs were reliably formed. If a kind of strong externalism about undercutting defeat is true (which I assume here for the sake of argument), this evidence is misleading in that it does not show that their way of arriving at their original beliefs was *modally* unreliable. But given the set-up of those cases, we have no reason to think that Larry or Anne (or the moral believer in the analogous moral

---

[111] Since even Lasonen-Aarnio thinks that there's something epistemically wrong with subjects who hold onto their beliefs in the relevant defeat-cases, this move also seems without independent support from any current work in epistemology. Baker-Hytch and Benton (2015) supplement (and go beyond) Lasonen-Aarnio's argument for the claim that knowledge can be retained in defeat-cases. But they also explain why subjects in these cases are epistemically irrational. Baker-Hytch and Benton argue that subjects in defeat-cases violate a derivative requirement of epistemic rationality: S must refrain from believing that p if S comes to believe or accept that one's belief that p is not knowledge. This requirement is derived from the so-called "knowledge norm on belief": S must not believe that p if S does not know that p. Subjects who persist in their beliefs in (undercutting) defeat-cases are violating this derivative norm, which is instrumental to fulfilling the primary epistemic norm of belief (p. 57):
"When one acquires evidence that renders it improbable that one knows that *p* such that one thereby comes to believe or accept that one doesn't know *p*, then by persisting in believing that *p*, one is violating the derivative norm…, even though one may in fact continue to know *p* (supposing one did know p to begin with)." (p. 58; italics in the original)
So, even the externalist defeat-sceptics (Lasonen-Aarnio and Baker-Hytch and Benton) at least agree that something is epistemically wrong with subjects who just persist in their beliefs in (undercutting) defeat-cases.

cases) are anything other than ordinary believers: i.e. subjects who are not able to reliably discriminate between

(i) cases, where the evidence in their possession is sufficient for making it the case that they should withhold or disbelieve that their relevant beliefs were reliably formed, but where this evidence is misleading,
(ii) from cases, where the evidence in their possession is not misleading.

For ordinary epistemic agents like these it is unreasonable to stick to their original belief in cases that belong to category (i), because it would foster a disposition in them to stick to their beliefs even in cases that belong to category (ii), where they are presented with non-misleading information or evidence. And this disposition would therefore fail to be knowledge-conducive over whole range of normal circumstances. Call the unreasonable disposition that is fostered here "dogmatism", i.e. a pattern of doxastic behaviour, that is overall not-knowledge-conducive, where a subject systematically sticks to beliefs in the face of or incoming information or evidence when, with the goal of knowledge in mind, she would be better served by revising her beliefs in most of those cases. Dogmatism here is basically the subjects' epistemically wrong disregard for incoming information or evidence that intuitively should prompt them to revise their beliefs.[112]

A subject who *dogmatically* sticks to her beliefs in the face of misleading evidence or information fosters an epistemically bad disposition in herself. This disposition is epistemically bad as the dogmatic subject will also be disposed to stick to her beliefs in cases where she is presented with non-misleading information or evidence. The subjects we are talking about here are not in a position to reliably discriminate between the two kinds of cases. To subjects like these, even in the second case, "it will seem

[112] I do not mean to deny that there are possible epistemic agents, that we can come up with, who have the ability to reliably discriminate between cases, where the relevant information or evidence is misleading, from cases where it isn't. As Lasonen-Aarnio is quick to point out, "[s]ince they make appeal to the dispositions and abilities that accompany the adoption of a method…evaluations [in terms of reasonableness] depend on the cognitive architecture of the subject under consideration" (ibid. p. 20). Subjects with the appropriate capacities would therefore perhaps not exhibit dogmatism in sticking to their beliefs, since it seems they do not run the danger of developing an overall non-knowledge-conducive disposition. But that should not come as any consolation to the standard responses, who (I assume) aim to protect the moral convictions of fairly ordinary, human epistemic agents.

to them as if they are following the same method as in good cases, thereby retaining knowledge, whereas they will in reality be retaining beliefs in falsehoods" (Lasonen-Aarnio 2010: p. 15). Overall, subjects should thus follow a belief-forming policy that recommends the revision of belief in the light of new information or evidence (ibid.), when this new information or evidence is such that it gives them a rationality undercutter for their original belief. This policy is also bound to result in the need to give up *modally* reliably formed true beliefs in some cases, as the examples of Larry and Anne illustrate, but overall, it is conducive to the attainment of knowledge in a wide range of normal cases. Overall, it seems better to be disposed to suspend judgment unless you can independently confirm the reliability of the relevant belief-forming faculty.

Assuming that knowledge is the goal of inquiry and following Cassam's (2016) account of intellectual vices,[113] we can call dispositions to form beliefs in a certain manner under certain circumstances that are not overall knowledge-conducive *vices* in virtue of their role "in impeding effective and responsible inquiry" (p. 169). Lasonen-Aarnio writes that epistemic reasonableness is "at least largely a matter of managing one's beliefs through the adoption of policies that are generally knowledge-conducive, thereby manifesting dispositions to know and avoid false belief across a wide range of normal cases" (Lasonen-Aarnio 2010: p. 2; cf. 12-17). Epistemic reasonableness is connected to knowledge because it constitutes a believer's ability to manage her beliefs through the adoption of knowledge-conducive policies.[114]

---

[113] Here's a passage nicely summarizing this account:
"What makes a character trait, thinking style or attitude intellectually vicious is its impact on our inquiries. Inquiry is the attempt 'to find things out, to extend our knowledge by carrying out investigations directed at answering questions, and to refine our knowledge by considering questions about things we currently hold true' (Hookway 1994: p. 211). In these terms, intellectual vices…can be characterized as intellectual character traits, thinking styles or attitudes that impede effective inquiry. Intellectual virtues, in contrast, are intellectual character traits, thinking styles or attitudes that abet effective inquiry. Examples might include open-mindedness, alertness, carefulness, and humility. An effective inquiry is one that is knowledge-conducive, and this casts light on why carelessness is an intellectual vice whereas carefulness is an intellectual virtue. Carefulness is knowledge-conducive whereas carelessness impedes our attempts to extend or refine our knowledge by inquiry." (Cassam 2015: p. 21)

[114] How are unreasonableness and intellectual vice connected? A subject is unreasonable when she adopts a non-knowledge-conducive belief-forming policy. In line with the account of intellectual vice offered by Cassam, I would then describe "unreasonableness" as an

Dogmatically sticking to your beliefs in light of counter-evidence or information sufficient for making it the case that you should withhold or disbelieve that your belief was reliably formed seems like a good example for exhibiting a belief-forming disposition that is intellectual vicious in Cassam's sense. It is intellectually vicious, because in believing as the dogmatic believer does, she exhibits a disposition for doxastic behaviour that is not-knowledge conducive over a whole range of normal cases, and thus in general both impedes effective and responsible inquiry and marks the believer as an untrustworthy source of information for others.

Assuming that Larry and Anne try to implement the recommendations respectively issued by the AMS and the third-factor accounts in their doxastic behaviour, they would retain their beliefs in light of the incoming information about what best explains their way of arriving at those beliefs. Since we now assume (for the sake of argument) that the Larry or Anne have a successful defeater-deflector at their disposal, their beliefs remain justified. But given everything we have said, it also seems to be the case that both Larry and Anne would exhibit a belief-forming disposition in their doxastic behaviour that is unreasonable for them to have, and that is therefore intellectually vicious. Given that the cases of Larry and Anne are in all relevant respects analogous to the cases of the relevant moral believers, it seems that for the respective moral believers it would also be intellectually vicious to hold onto their beliefs – even if their moral beliefs remain justified in light of the considerations powering the generic EDA. Therefore, even opting for the defeater-deflector-option, and working on the assumption (for the sake of argument) that the standard responses are successful in protecting the justification of our moral beliefs via preventing the occurrence of defeat, the standard responses can hardly be called a success.

The important lesson to take from the discussion of the cases of Larry and Anne in this section is that dogmatically sticking to your original belief and ignoring incoming counter-evidence (or apparent counter-evidence) is epistemically wrong even in cases where your belief was *modally* reliably formed and where you justifiably belief that it is *modally* reliable, if true. It seems that being disposed to simply ignore incoming counter-evidence in good cases would also dispose you to simply ignore incoming

---

umbrella term, that broadly groups together a variety of intellectually vicious forms of doxastic behaviour (i.e., forming and retaining beliefs and collecting and assessing evidence).

counter-evidence in bad cases that look like good cases from your vantage point. Overall, it would be better to be disposed to suspend judgment unless you can independently confirm the reliability of the relevant belief-forming faculty.

4.7 THE DOGMATISM ARGUMENT

What now remains to be done is to apply this lesson to the moral case. I will do so by explicitly stating the argument I have developed in this section. This argument presupposes several things. First and foremost, I assume the following:

> ***DEFEATER-DEFLECTOR ASSUMPTION***: The standard responses try to prevent the occurrence of defeat.

In the argument below, I assume that the standard responses are successful in showing that if our moral beliefs are true, then they are *modally* reliably formed. I assume that our moral beliefs are true. And, for the sake of argument, I assume that the presence of a rationality undercutter does not suffice to render a belief unjustified, if this undercutter does not show that our moral beliefs were not *modally* reliably formed. What the following argument, which I have termed the "Dogmatism Argument", then shows is that even assuming that the standard responses are successful in preventing the occurrence of defeat, it is nonetheless the case that holding onto our moral beliefs in the presence of information DEBUNKING (cf. section (3.5)) is epistemically wrong.

And now here's the *Dogmatism Argument*: Upon being confronted with information DEBUNKING at t_2, we have a rationality undercutter for our moral beliefs. We continue to hold onto our moral beliefs after t_2, as the standard responses recommend, since we see that if our moral beliefs are true, then they are *modally* reliable. This assumes that we apply the reasoning urged on us by the standard responses to our case and hold onto our moral beliefs: we see that DEBUNKING does not give us a reason to think that our beliefs are not *modally* reliable, if they are true. And we simply assume that our moral beliefs are true. But, as I have argued above, information DEBUNKING nonetheless seems sufficient to generate a rationality undercutter for our moral beliefs. This premise is supported by the claim that subjects like Larry and Anne, whose cases are analogous to the cases of the relevant moral believers have a rationality undercutter for their respective beliefs.

Nothing else happens at or after t_2, that counter-acts the rationality undercutter in our possession. This assumes that nothing else happens that would make it the case that we lose the relevant rationality undercutter.

And now the real action starts: assume that you are epistemically entitled to disregard a rationality undercutter regarding a belief or a set of beliefs of yours only if you are in a position to appreciate the fact that this evidence is misleading and to reliably discriminate between this case and similar cases, where a similar rationality undercutter is sufficient for defeating the justification of your belief. This premise spells out a necessary condition that must be fulfilled for a subject to be entitled to disregard a piece of information that constitutes a rationality undercutter.

I have argued that subjects are not entitled to just disregard rationality undercutters in their possession, even in cases where this rationality undercutter does not actually take away their justification (since we assume a strong externalism about defeat). Subjects are not entitled to do so, since this doxastic behaviour fosters in them a disposition that is not knowledge conducive over a large range of similar cases, and that is therefore intellectually vicious.

We, as ordinary moral believers, are not in a position to appreciate the fact that DEBUNKING is misleading and to reliably discriminate between this case and similar cases, where a similar rationality undercutter is sufficient for defeating the justification of your belief. Therefore, we are not entitled to disregard this rationality undercutter regarding our moral beliefs.

I term unentitled disregard "dogmatic ignorance". Displaying dogmatic ignorance renders subjects vicious in exhibiting and fostering in them a non-knowledge conducive disposition for retaining beliefs and evaluating evidence. Dogmatically ignoring (=unentitled disregard for) a rationality undercutter regarding a set of your beliefs renders you epistemically vicious in holding onto your relevant beliefs.

It then follows that we, as ordinary moral believers, who fail to satisfy the condition for entitled disregard of incoming information, are dogmatically ignoring a rationality undercutter regarding our moral beliefs. So, in holding onto our moral beliefs after t_2, we are epistemically vicious. We are exhibiting epistemic vice in holding onto our moral beliefs after being confronted with DEBUNKING at t_2.

What the Dogmatism Argument shows is that even if the standard responses were successful in protecting the justification of our moral beliefs, it would nonetheless be epistemically wrong for us to hold onto our moral beliefs in the presence of DEBUNKING. What this shows is that even the best-case scenario for the standard responses (where they succeed in protecting the justification of our moral beliefs) is still pretty bad, to say the least. This concludes my line of argument in this chapter.

5 CONCLUSION

By way of conclusion, let me now try to put together the most important points I have made in chapters (3) and (4). The winding path we have taken in critically discussing the AMS and third-factor accounts now puts us into a position to sum up the crucial insights gathered in our critical discussion in the form of a simple and straightforward argument.

First, assume a plausible desideratum on any response to the generic EDA (or arguments like it): any satisfying response should support the claim that evolutionary considerations do not show that our moral beliefs are seriously epistemically deficient or that we are seriously epistemically deficient for holding them. In brief: a good response to the generic EDA should show that evolutionary considerations do not suffice to render us epistemically criticisable for holding or continuing to hold our moral beliefs.

With this desideratum in hand, the following argument now unfolds: in response to the generic EDA, the standard responses are either to be regarded as *defeater-defeaters* or as *defeater-deflectors*. If the standard responses are regarded as *defeater-defeaters*, then they are unsuccessful in reinstating the justification of our moral beliefs. So, the first horn of this dilemma is a complete non-starter.

If the standard responses are regarded as *defeater-deflectors*, then it is nonetheless the case that we would be epistemically vicious to hold onto our moral beliefs after we were confronted with DEBUNKING. As we have seen in the last section, even getting to the point that we are *only* epistemically vicious in continuing to hold our moral beliefs comes with serious epistemic baggage, as it commits the AMS and the third-factor accounts to a quite strong epistemic externalism about justification and defeat. Even then, evolutionary considerations suffice to make it the case that we would be epistemically criticisable for continuing to hold our beliefs.

So, I conclude that the AMS and third-factor accounts are bad responses to the generic EDA. They are bad responses, as they fail to show that evolutionary considerations do not render us epistemically criticisable for holding or continuing to hold our moral beliefs. This is the central conclusion of this thesis.

I should make clear that the intended result of my arguments is *not* that an argument along the lines of the generic EDA is successful. The result I have argued for is that arguments like the generic EDA are not unsuccessful *on account of the reasons provided by the standard responses*. Yet these debunking arguments might fail for other reasons.[115]

What I hope to have shown is that we have good reason to believe that certain popular ways of responding to EDAs on behalf of NNMR are (epistemologically speaking) deeply flawed. This result suggests that the popularity of these responses is quite unearned, at least in the context of the debate about evolutionary debunking in moral epistemology, and that this debate should be re-oriented towards paying greater attention to other issues with EDAs.

[115] Cf. appendix C for some empirically-minded responses to EDAs.

6 APPENDICES

6.1 APPENDIX A: INTERNALISM & EXTERNALISM ABOUT EPISTEMIC JUSTIFICATION & MENTAL STATE DEFEATERS

In this appendix, I will provide a rough characterization of internalism and externalism about epistemic justification, and I will state why I think that the framework of undercutting mental state-defeat presented in section (1.4.1) does not beg the question against (at least many versions of) externalism.[116]

The debate (or rather debates) between internalism vs. externalism in epistemology is very roughly about whether certain necessary conditions for the possession of some positive epistemic status (e.g., justification, the possession of evidence, knowledge) are "internal" to the believer in a relevant sense (Bergmann 1997: p. 399). In which sense depends on the positive epistemic status in question, and on the forms of internalism and externalism contrasted with each other, e.g., access internalism holds that if a person has knowledge or justified belief, she has or could have access to the basis of her knowledge or justified belief (where "access" could be cashed out in terms of "knowability by some introspective or reflective method" (Goldman & Beddor 2015: section 2)). Externalists will deny that access in this sense is required for knowledge or justification.

For our purposes here, all we need is a rather basic idea about one point of contention between two varieties of epistemic internalism and externalism. In this thesis, I will understand "internalism" and "externalism" as two conflicting approaches towards epistemic justification. The conflict between internalism and externalism about epistemic justification arises from the fact that these views (or families of views) disagree about whether it is necessary for S's belief that p to be justified that S has or could have access (in the sense of knowability through reflection) to (at least some essential part of) the *justifier* for S's belief that p. "Justifiers" are those things, which

[116] There are epistemic externalists who are sceptical about the framework of epistemic defeat, and so, it is no wonder that they would object to NO DEFEATER, cf. section (4.5). Here, I just want to show that there are versions of epistemic externalism, who should be perfectly compatible with NO DEFEATER.

for any "… given justified belief… make up or constitute the person's justification for that belief at that time" (Pappas 2014: section 3).

Internalists about epistemic justification now assert that potential knowability through reflection of at least some essential part of the justifier for S's belief that p is necessary for S's belief that p to be justified. Externalists deny that this is a necessary condition for epistemic justification as such. Therefore, externalists in this sense could e.g., allow for S's belief that p to be justified because her belief was produced by a reliable faculty in appropriate circumstances – even though the fact that her belief was produced by a reliable faculty in appropriate circumstances is not reflectively accessible to S.

With this sketch of the distinction between internalism and externalism out of the way, let us next consider the following question:

> Does the focus on "mental state" defeaters in my thesis bias the account of defeat (presented in section (1.4.1)) towards epistemic internalism?

No. First, Bergmann, who defends a version of NO DEFEATER himself, is an epistemic externalist about justification. And prominent epistemic externalists apart from Bergmann acknowledge the importance of mental state defeaters (cf. Nozick 1981: p. 196; Goldman 1986: pp. 62–63, pp. 111–112).

Secondly, it is not obvious why these commitments should be problematic for an externalist in the above sense: NO DEFEATER is perfectly compatible with S's belief being justified and S never having considered whether her belief is formed in a trustworthy fashion. Furthermore, acknowledging the importance of mental state-defeaters seems to be easy enough for most externalists, since this only requires conceding that a subject's mental states can make *some difference* to the justificatory status of a belief of that subject. Conceding this point does not obviously require you to endorse a claim that clearly conflicts with epistemic externalism about justification (such as e.g., that *only* a subject's mental states are relevant for the justificatory status of that subject's beliefs).

Thirdly, it is independently plausible that it reflects negatively on the epistemic status of your belief that p if you should withhold or disbelieve that your belief that p was formed reliably, given everything else you believe or should believe. And so, epistemic externalists at any rate should try to account for this intuitive thought.

6.2 APPENDIX B: SENSITIVITY AND SAFETY

As stated in section (1.3.5), modal conditions are often used in epistemology to cash out the intuitive idea that knowledge requires our beliefs to be not merely correct, but to stably track the truth even if your circumstances would slightly change (Ishikawa & Steup 2017: section 5). The hope is that by applying these modal conditions, we can filter out cases that involve epistemic luck of the kind that is incompatible with knowledge. So, here's a first stab at sensitivity and safety:

> ***Sensitivity$_1$.*** S's belief that p is sensitive if and only if: if p were false, S would not believe that p.
>
> ***Safety$_1$.*** S's belief that p is safe if and only if: if S were to believe that p, p would be true.[117, 118, 119]

On Lewis' (1973) semantics for counterfactual conditionals, the sensitivity condition is equivalent to the requirement that in the nearest possible world(s) in which not-p, S does not believe that p (Ishikawa & Steup 2017: section 5.1). The safety condition is then equivalent to the requirement that in the nearest possible world(s) where S believes that p, p is true.[120]

---

[117] Cf. Sosa (1999): "A belief is sensitive iff had it been false, S would not have held it, whereas a belief is safe iff S would not have held it without it being true. For short: S's belief B(p) is sensitive iff $\neg p \Box\!\!\rightarrow \neg B(p)$, whereas S's belief is safe iff $B(p) \Box\!\!\rightarrow p$." (p. 146)

[118] As Sosa (1999) points out, although contraposition is valid for material conditionals (i.e., "If A then B" is logically equivalent with "If not-B, then not-A") it is not for counterfactuals, and so sensitivity and safety are not equivalent. To see this, consider the two propositions (α) "If John had gone to the party, Joanna still would have gone to the party" and (β) "If Joanna had not gone to the party, John still would not have gone to the party" (cf. Lewis 1973: p. 35). It should be clear that (α) is not equivalent to (β) since even assuming (α) is true, (β) might still be false, as John might still have gone to the party had Joanna been absent (cf. Rabinowitz 2011).

[119] Apart from the point made in FN 118, it is also worth to emphasize a point nicely summed up by Comesaña in the following passage:

"A subjunctive conditional does not have the same truth-conditions as a strict conditional, for a subjunctive conditional can be true even if there are some worlds where the antecedent is true and the consequent false, provided that those worlds are different enough from the actual world (for instance 'If Mike Tyson were to fight David Letterman, then Mike Tyson would win' is certainly true, even though there surely are possible worlds where Letterman wins)." (2007: p. 782)

[120] There is a technical complication that I gloss over here, but that I want to flag: As many commentators point out (Ichikawa 2011: p. 309, FN 22; Ichikawa & Steup 2017: section 5.2; Comesaña 2007: pp. 786-787; Roland & Cogburn 2011: p. 553), understanding safety in terms of counterfactuals does not seem to work well with the standard semantics for

But what's the notion of "nearness" of or "closeness" between worlds involved here? As Hawthorne (2004: p. 56) notes (when discussing the safety-condition, but the point also applies to the sensitivity-condition) the kind of closeness that is of interest to epistemologists when they propose their modal conditions cannot be cashed out in terms of any general-purpose notion of metaphysical closeness at play in the discussion of counterfactuals.[121]

Rabinowitz writes that for epistemic purposes which world counts as relevantly similar to the world w in which a subject S believes p at time t via method M is to be determined on a case by case-basis mostly (but not exclusively) relative to "the conditions of belief-formation" represented by the set of the belief that p, the time t, the subject S, and the method M by which S formed the belief p at t in w (2011: section 3). (The question of relevant similarity is also relevant in the discussion of the AMS, cf. section (3.2)). I would now like to focus on one of these conditions of belief formation, the method by which S formed her belief.

When assessing a belief for safety or for sensitivity, it is important to hold fixed the way in which the agent arrives at or sustains her belief (i.e., the subject's method or basis for belief), as otherwise applying these modal conditions won't serve to capture the anti-luck intuitions they are intended to capture (Roland & Cogburn 2011: p.

---

counterfactuals. Lewis' account of counterfactuals includes a so called "centring condition" according to which the actual world is always the uniquely closest world. This would make safety a trivial condition, as all true beliefs would then trivially count as safe, because all actually true beliefs are also true in the next closest possible world if the next closest possible world is the actual world. (Ichikawa (2011: p. 309) sketches out an option for safety-theorists like Sosa (1999). In the same paper, Ichikawa also offers a semantics for counterfactuals that avoids this problem. But on this semantics, sensitivity and safety are equivalent.)
In light of this issue, Ichikawa and Steup propose that it "may be most perspicuous to understand the safety condition more directly in…modal terms, as Sosa himself often does" (2017: section 5.2).

[121] According to David Lewis' (1973) account, a counterfactual is true just in case the corresponding material conditional holds in all the worlds in a relevant set. The relevant set of worlds is the set of worlds where the antecedent is true that are more similar to the actual world than any worlds where the antecedent is false. So, to evaluate if p were the case, then q would be the case, we examine the sphere of p-worlds that differs less than any non-p-worlds from the actual world and check whether q is true in those worlds. Here we can see how Lewis' metric for similarity and similarity for epistemic purposes might come apart. On Lewis' account a world w* can for instance fail to be relevantly similar to the actual world w if it has a different history. But as Rabinowitz (2011: section 3) points out, when it comes to safety, possible worlds with a different history to w can nevertheless count as close, provided that the conditions for belief-formation are the same or similar in both worlds.

550).[122] Consider first cases relevant to the sensitivity-condition, where a subject believes that p on good grounds or by using a good method, but S could believe that p on bad grounds or by using a bad method, even if p were not the case (Ichikawa 2011: p. 302). Consider Ichikawa's example from the same passage:

> A wife believes, on the basis of excellent evidence, the truth that her husband is faithful. But she is psychologically unable to face difficult truths. So, had her husband been unfaithful, she would have self-deceived herself into irrationally believing him to be faithful. (ibid.)

Here we can see the need to index the sensitivity-condition to the actual way in which the subject acquires her belief: ***Sensitivity$_1$*** gives you the result that the wife's true belief acquired through considering excellent evidence is not sensitive and therefore does not constitute knowledge. But that's counter-intuitive: the wife believes truly that her husband is faithful due to top-notch evidence.

The analogous point also applies to safety: Consider a similar case, where the wife again actually responds properly to the evidence in her possession, but her belief fails to satisfy ***Safety$_1$*** (ibid.: p. 309):

> A wife believes, on the basis of excellent evidence, the truth that her husband is faithful. But she is psychologically unable to face difficult truths. Moreover, her husband might have easily been unfaithful. In that case, she would have believed that her husband is faithful via self-deception, even though this belief would be false in that case.

Safety$_1$ gives you the result that the wife's true belief acquired through excellent evidence is not safe and therefore does not constitute knowledge. Again, this result seems counter-intuitive. Therefore, if the two modal conditions are not relativized to methods, they do not track the salient instances of epistemic luck.[123]

Therefore, we need to relativize to the subject's methods or basis for belief to avoid these problems. Here are amended versions of the two modal conditions: let "M"

---

[122] Roland and Cogburn call this requirement the "Constancy-Principle":
"When considering changes in an agent's doxastic attitude with respect to a proposition across worlds, the process/method by which the agent acquires (or sustains) that attitude must remain constant." (2011: p. 550)

[123] I am side-stepping the worry here of how we ought to individuate methods (cf. Rabinowitz 2011: sec. 3.1.3).

denote the way in which the subject S acquires the belief that p, i.e. the subject's basis for belief or her method for attaining it:

> ***Sensitivity₂***. S's belief that p formed via or based on M is sensitive if and only if: if p were false, S would not believe that p via or based on M.
>
> ***Safety₂***. S's belief that p formed via or based on M is safe if and only if: if S were to believe that p via or based on M, p would be true.

In order to leave the problem flagged in FN 120 aside, I follow Ichikawa's and Steup's advice (cf. FN 120 above):

> ***Sensitivity₃***. S's belief that p formed via or based on M is sensitive if and only if: in nearby possible worlds where p is false, S would not believe that p via or based on M.
>
> ***Safety₃***. S's belief that p formed via or based on M is safe if and only if: in nearby possible worlds where S believes that p via or based on M, p is true.[124]

A few examples might help to elucidate the two modal conditions a bit further. Robert Nozick's (1981) influential statement of the sensitivity-condition was motivated by the anti-sceptical potential of the condition. According to ***Sensitivity₃***, my true belief that there is a cherry tree in my garden (formed via perception) can count as knowledge, since if there were no cherry tree in my garden, I would not believe that there is a cherry tree in my garden via perception. In the close possible worlds in which there is no cherry tree in my garden (e.g., where none is planted there; where I had it cut down), I do not believe that there is a cherry tree in my garden via perception.

Sceptical scenarios do not impinge on my knowledge, as worlds where these scenarios apply are more dissimilar to the actual world than the world where no cherry tree is planted in my garden and are therefore "further off" (Rabinowitz 2011: section 1).

---

[124] There are several complications I am glossing over, but I want to flag one of them in this footnote. As Timothy Williamson (2000: p. 100) points out (with respect to safety, but the point also applies to sensitivity), the modal conditions are vague with respect to the question of "What counts as a close world?" because the notions meant to be elucidated by the modal conditions (i.e., "relevant similarity", "reliability" and "knowledge") are also vague. Due to this vagueness, Williamson argues that there will be cases in which whether you think that there is a close world in which the agent falsely believes depends on whether you are willing to attribute knowledge or reliability to the agent in the actual world in that case (ibid.; Williamson 2009: pp. 305ff).

This seems plausible. Radical sceptical scenarios typically involve quite fantastical elements or scenarios like e.g., the activity of powerful Cartesian demons. It is intuitive to think that worlds where these sceptical elements or scenarios obtain are highly dissimilar to the actual world. Thus, the fact that I would have false beliefs in the sceptical worlds[125] is irrelevant with respect to the question of whether my belief amounts to knowledge in the actual world.

Although originally motivated on account of its anti-sceptical prowess, sensitivity has also been criticised for struggling with "abominable conjunctions" in those kinds of scenarios (cf. DeRose 1995: pp. 27–28). Let's suppose that scepticism with respect to ordinary knowledge is false. Let's further suppose that I believe that I have hands (as I do). Given ***Sensitivity$_3$***, and everything else being equal,[126] I know that I have hands, since in nearby worlds where I do not have hands (e.g., because there I recently had a horrible accident), I would not believe so via the same method or on the same basis. Now it seems incredibly intuitive to hold that if I know that I have hands, and I know that if I have hands this entails that I am not a handless brain in a vat (which seems to be a plausible assumption), then I know that I am not a handless brain in a vat.[127]

However, on the sensitivity-account, I fail to know that I am not a handless brain in a vat. The reason for this is that I would falsely believe that I am not a handless brain in a vat in the closest world in which the proposition "I am not a handless brain in the vat" is false, i.e. the world in which I am a handless brain in the vat (Rabinowitz 2011). This leads to an *abominable conjunction*: I know that I have hands and I do not know that I am not a handless brain in a vat (DeRose 1995: pp. 27-28). This is one of the reasons why many epistemologists today reject sensitivity as a necessary condition for knowledge, but nonetheless there are also defenders of the condition still standing (cf. several of the essays in Becker & Black 2012; Ichikawa 2011).

Safety-accounts have been explicitly motivated by their ability to avoid these conjunctions. Consider for instance Sosa's (1999) example of a belief that is safe, but

---

[125] In those cases it is part of the sceptical hypothesis that the relevant truths would have been different while my beliefs would have been the same.

[126] That is to say, if all other things necessary for knowledge are in place.

[127] The principle underlying this intuitive thought is that knowledge is closed under entailment (often referred to as the "Closure-Principle"), and roughly holds that if S knows p, and if S knows that p entails q, then S knows q.

not sensitive, i.e. the belief that a distant sceptical scenario does not obtain (e.g., "I am not a handless brain in a vat"). If we stipulate again, that scepticism regarding ordinary knowledge is false, then the belief that I am not a handless brain in a vat is safe, although it is not sensitive, as we have just seen.[128]

To see why, consider this case: Let's say that we have a subject here called "GE Moore". GE Moore forms the true perceptual belief that "I have hands". Given ***Safety$_3$***, Moore knows that he has hands, since in nearby possible worlds where he also forms the belief that he has hands via perception, he has hands.

He then makes the following inference: "If I have hands, then I am not a handless brain in a vat". Again, it seems plausible to say that if Moore knows that he has hands, then he knows that if he has hands, he is not a handless brain in a vat. His belief in the conditional "If I have hands, then I am not a handless brain in a vat" is also safe: in nearby worlds where Moore believes in this conditional, it is true that if he has hands, then he is not a handless brain in a vat.

So, he infers from his perceptual belief the claim that "I am not a handless brain in a vat", which he then believes. GE Moore's true inferred belief "I am not a handless brain in a vat" is safe, since even if the circumstances surrounding Moore were to slightly change, his belief formed via inference would then still have been true. In nearby possible worlds, where Moore gains the belief that he is not a brain in a vat via the above inference, this inferential belief is still true.

Moreover, the following point applies again: the sceptical worlds where Moore has the inferential belief and where it is false are more distant and are therefore not relevant. Therefore, Moore can know that he has hands, that if he has hands, then he's not a handless brain in a vat, and that he is not a handless brain in a vat.

---

[128] Here's Sosa's own short summary on why the belief that a sceptical scenario does not obtain is safe:
"[N]ot easily would one believe that not-H [where H is the claim that a sceptical scenario obtains] (that one was not so radically deceived) without it being true (which is not to say that not possibly could one believe that not-H without it being true). In the actual world, and for quite a distance away from the actual world, up to quite remote possible worlds, our belief that we are not radically deceived matches the fact as to whether we are or are not radically deceived". (Sosa 1999: p. 147)

And safety-accounts also seem to work fine in non-sceptical cases: My true perceptual belief that there is a cherry tree in my garden can count as knowledge, since in the next closest worlds where I have this belief, but where things are slightly different, the belief is still true. Moreover, in the relevant nearby worlds where I do not have this perceptual belief, this is so because in those worlds there is no cherry tree in my garden for me to perceive.[129, 130]

It should be added that not everyone agrees that either safety or sensitivity are necessary for knowledge. For purported examples of insensitive knowledge, cf. Vogel (2012). For purported examples of unsafe knowledge, cf. e.g., Bogardus (2014). Although I do find these examples convincing, my argument in this thesis does not depend on their success.

6.3 APPENDIX C: IDEALIZATIONS

In section (1.4), I try to investigate whether we can find an epistemologically respectable epistemic challenge to our moral beliefs that arises from making the discovery that evolution has influenced *our* (i.e., *your* and *my*) moral beliefs in some way. From (1.4), and throughout the rest of my thesis, I assume that we indeed have good reason to believe that evolution has influenced our (i.e., your and my) moral beliefs in a way that is apt to give rise to the challenge that is formulated by the generic EDA. But this assumption presupposes several idealizations that need to be made explicit.

First off, it assumes that there is a considerable evolutionary influence on the moral convictions of contemporary humans, and that evolution (perhaps combined with other causal factors, like e.g., cultural influences) explains why contemporary humans have (at least many of) the moral convictions they do have. Some critics of EDAs dispute this. For example, Jessica Isserow (2018) argues that we are not in a position

---

[129] In many cases safety and sensitivity converge, i.e. in many cases, if your belief is safe, then it is also sensitive. This is indeed plausibly the case in the example above, where it is also true to say, that in nearby worlds, where there is no cherry tree in my garden (e.g., because I had it cut down), I do not have the perceptual belief that there is a cherry tree in my garden.

[130] When it comes to cases concerning knowledge of the negation of sceptical hypotheses, the safety-condition is less demanding than the sensitivity principle. However, there are also possible cases where safety is more demanding than sensitivity, cf. Rabinowitz (2011: section 3b).

to place a sufficiently high level of confidence in any account of our moral evolution. Therefore, typical EDAs fall short of generating a defeater. And William Fitzpatrick (2015; 2016) argues, that in claiming that our moral beliefs across the board are nothing but the results of truth-indifferent processes, the debunkers simply overreach. According to him, what the debunkers fail to take into account is the possibility that although the psychological capacities we utilize in forming moral beliefs evolved for Darwinian reasons, we may nonetheless be able to develop our abilities for moral reflection and reasoning in cultural contexts and exercise them with a considerable degree of independence from those truth-indifferent influences (Fitzpatrick 2016: p. 437, cf. Fitzpatrick 2015).

Furthermore, the above assumption presupposes something more than just that evolutionary forces have influenced the moral beliefs of contemporary humans. It assumes that evolution has influenced *your* moral beliefs and *my* moral beliefs. Evolution (it is assumed) accounts for why individual contemporary humans have many of the moral beliefs they do have. However, this assumption is also controversial in its own right. Mogensen (2016b) and Hanson (2017) have each pointed out that this assumption seems to conflict with an influential view in the philosophy of biology, the view that natural selection only explains frequencies in which traits occur in populations (cf. Sober 1984). Since this view holds that natural selection cannot explain the traits of individuals, it seems to entail that the facts about past selection cannot form part of a discrediting explanation of the moral beliefs of any individual (Mogensen 2016b: pp. 1805-1806).

These are serious worries, but I am going to set them aside in this thesis. By setting them aside, I do not mean to detract from their importance. Rather, I am bracketing these worries to make it easier to consider in abstract terms whether an evolutionary moral genealogy could ever amount to or be part of a coherent epistemic challenge (Locke 2014: p. 223) – *despite* the considerations offered by the standard responses. This is the topic of my current undertaking. Once you have explored this possibility, you can still take away my assumptions and consider whether the differences between the idealized and a more realistic scenario make an epistemically relevant difference.

The topic of this thesis is whether certain standard responses to EDAs are successful. Obviously, I think it is a fair response to EDAs to call into doubt the relevant

empirical assumption involved in those arguments. But importantly, this route is not taken by these standard responses, and indeed, it seems that the empirical responses, if successful, would render the standard responses superfluous: if EDAs fail for reasons related to their empirical assumptions, why would we need to engage in the quite intricate manoeuvres involved in the standard responses (cf. also section (3.3.2), FN 86)?[131]

This point motivates several idealizing assumptions that I will make concerning evolution and human moral psychology. I take these useful idealizing assumptions concerning the evolutionary impact on our moral beliefs from Dustin Locke (2014: pp. 222-223) although I have slightly amended the terminology:[132]

(a). All of our moral beliefs are directly based on our *moral faculty*.

(b). Our moral faculty has a *direct* and *complete* Darwinian explanation. This explanation holds that in the environment of evolutionary adaptation, our ancestors' moral faculty was naturally selected to produce adaptive moral beliefs, and this is why we now have this faculty.

(c). There is no epistemically relevant difference between our *current environment* and the *environment of evolutionary adaptation*.

Why make these three assumptions? These assumptions abstract away from the far messier and more complex picture of our actual moral psychology and its evolutionary history. It is precisely the messiness and complexity of the actual picture, which motivates criticisms in the vein of Fitzpatrick, Hanson, Isserow and Mogensen, and

---

[131] This harkens back to a point I have made in the introduction (chapter 0) to this thesis.

[132] Instead of speaking of "moral dispositions" as Locke does, I talk of the "moral faculty". I am not sure if there is more than a terminological difference between Locke and me here, but I will of course provide my rationale. I use the term "faculty" interchangeably with various terms: "methods", "mechanisms", "sources" and "processes". These all refer to the same thing, i.e. our way of arriving or forming specific beliefs. I use the term "belief-forming dispositions" in a way that is not equivalent to the above terms. Perception, introspection, deductive and inductive inference are all faculties/mechanisms/methods/processes/ sources – but e.g. being open-minded, being diligent in gathering and evaluating evidence or being disposed to dogmatically refuse to take new evidence into account are not. The latter examples are what I will call "belief-forming dispositions". These are traits of the intellectual character of a person, that influence the belief-forming and evidence-gathering habits of that person. For a similar distinction between ways of arriving at specific beliefs and intellectual character traits cf. Lepock 2011.

which I want to idealize away from to inquire into the epistemological underpinning of debunking arguments.[133]

Assumption (a) holds that our moral beliefs are formed via the workings of a moral faculty. One way of thinking about the relationship between our moral beliefs and the moral faculty is to hold that our moral beliefs "depend crucially on intermediate conscious states, delivered by our moral faculty" (Bogardus 2016: p. 641). The term "moral faculty" then just refers to whatever produces these intermediate conscious states. Different views on moral psychology will fill in different intermediate conscious states here: moral sentiments, gut reactions, affect-laden intuitions, intellectual seemings, pure and affectless conscious thoughts etc. Our moral faculty produces the intermediate conscious state and this state then regularly and directly prompts the formation of moral beliefs. Tomas Bogardus calls this popular family of views "Representationalism" (ibid.).[134]

Condition (a) idealizes first of all away from the fact that many of our moral beliefs are not simply based on the workings of this moral faculty, but are e.g., the product of chains of reasoning, of testimony etc. Fitzpatrick's critical point above is thus bracketed. Secondly, since (a) also entails that the moral beliefs of every individual human moral believer are directly based on the moral faculty, assuming (a) also allows us to bracket Mogensen's and Hanson's worry. And thirdly, (a) leaves aside other accounts of moral psychology, like e.g., rationalism or divine command theory (Bogardus 2016: p. 642).

---

[133] This epistemological underpinning might still be of interest even if it turns out that evolutionary debunking fails once we remove these idealizing assumptions (cf. Locke 2014; Lutz 2017; Vavova 2014: p. 79). We may well be able to apply the abstract epistemological story (developed in section 1.4, summarized in appendix I) to other cases in which a set of beliefs was apparently formed under the influence of epistemically-irrelevant factors.

[134] Bogardus (2016) emphasizes that moral psychology is of key importance in determining the scope of EDAs. Assume that an EDA can be found that rests on a sound epistemological principle. Bogardus argues that even then, this EDA will only be of danger to moral non-naturalists who accept Representationalism. Both rationalists who hold that we can gain direct access to the moral truth via rational reflection and theists who hold that some of their moral beliefs are in one way or another ultimately produced by divine activity can rationally maintain that their justification for such beliefs is not threatened by even the most challenging kind of EDA. He therefore concludes that only those who accept a broadly *naturalistic* worldview should be concerned by EDAs.

Assumption (b) idealizes away from the fact that the actual empirical explanation of our moral belief-forming processes will likely be very complex (i.e., not solely consisting of evolutionary factors) and far less direct. As Matt Lutz (2017: p. 13) writes, the appeal of evolutionary explanations to the debunker is that they complement sociological and neuropsychological accounts of human moral conviction and thus seem to go a long way toward completing a naturalistic explanation of our moral beliefs that makes no mention of the mind-independent, non-natural moral truth. (b) simplifies the actual situation by stating that such a complete naturalistic explanation is provided by an available evolutionary account.

I also think it is true that as things currently stand, we should not be too confident about any one explanation of our moral evolution. This is precisely Isserow's point. But for the purposes of formulating a generic EDA, I will assume below (in section (1.4.5): see assumptions (I) and (II) stated there), that the above direct and complete Darwinian explanation is among all the respectable available explanations of our moral faculty the one that is clearly best supported by the overall evidence. It should be added, that my assumption is not that this Darwinian explanation is so exceedingly well supported that it is hard to even begin to raise doubts about it in a rational fashion. Rather, I assume that the relevant direct and complete Darwinian explanation is sufficiently supported to constitute both (i) a sufficiently good explanation of our moral faculty, and (ii) the clearly best explanation currently available to "us" (i.e., everyone who gets informed about the current state of research on the evolution of morality).[135]

Assumption (c) enables us to draw direct inferences from the reliability of the moral faculty in our ancestors' circumstances to the reliability of the moral faculty in our circumstances. This enables us to draw inferences like e.g., if the moral faculty reliably produced true beliefs in the environment of evolutionary adaptations, then it still reliably produces true beliefs. This is important, as whether a certain belief-forming mechanism is reliable or not depends in part on the circumstances in which it is used.

---

[135] It is important to mention these two points separately, as it is possible for an explanation to be the best explanation available to us, while this explanation is still exceedingly poor. I also assume here that part of what makes an explanation "good" is that it is well supported by the available evidence.

Even perception, the prime example of a typically highly reliable mechanism, will not produce accurate beliefs in a stable and trustworthy manner in all kinds of circumstances. You should not trust your sense perception to the same extent as you would in normal circumstances when it is e.g., dark and foggy outside, or when the distorting effects of heavy medication affect you. (c) prevents the many differences between our current circumstances and the circumstances of evolutionary adaptation from making an epistemically relevant difference of this sort.

No doubt, my idealizing assumptions certainly make things easier for the aspiring debunker. But they do not make things too easy: the two standard responses are not impeded by my idealizations.[136] What these idealizing assumptions in effect accomplish, is that the generic explanatory EDA does not succumb to the empirically-minded responses to EDAs, which are independent of the standard responses. These idealizations make sure that there is a real job to do for the standard responses. As I have stated in the introduction, the allure of these responses is that they promise to neutralize the epistemic threat supposedly arising from evolution, even under the assumption that the debunkers' empirical hypotheses are largely correct, and if we assume NNMR.

6.4 APPENDIX D: RELIABILITY

In this appendix, I basically just want to give you quick idea of two popular and distinct uses of the term "reliability" in contemporary epistemology.

According to one established sense of the term, "reliability" refers to the sort of modal stability of the connection between belief and truth that characteristically appears to be missing in Gettier-cases, where the relevant subject ends up with a true belief but does so in a manner which is problematically accidental (cf. the example of Bertrand in section 1.3.5). This sort of modal reliability is cashed out by the modal conditions of knowledge that we have already come to know.

A different understanding of the term "reliability", which can come apart from modal reliability (cf. Baker-Hytch 2014: pp. 175-176) refers to "process reliability" (cf.

[136] As I have mentioned in the introduction, for the standard responses to be non-superfluous, we need to assume that EDAs are somewhat resistant to other empirical and epistemological criticisms. Otherwise, it is hard to see what the appeal of these responses is.

Becker 2008; Pritchard 2012a): i.e., "the salient process-type that produced… [the subject's] belief has a high truth-ratio in the actual world and in worlds with similar physical laws" (Baker-Hytch 2014: p. 176).[137]

To see that modal reliability and process reliability can come apart, consider that a false belief cannot be reliably formed, if we understand the notion in terms of *modal reliability*. The modal connection between belief and truth cannot be sufficiently modally stable (for knowledge), if it came apart in the actual world! This is reflected in the safety-condition and the sensitivity-condition, which both presuppose that the relevant belief is true in the actual world.

On the other hand, if we suppose that the term "reliability" refers to *process reliability*, then a belief can be reliable, even though it is false, on account of being produced by a salient process type that has high truth-ratio in the actual world and in worlds with similar physical laws. Having a high-truth ratio is compatible with producing some false beliefs.

6.5 APPENDIX E: NO DEFEATER

Here I want to provide some clarifications regarding NO DEFEATER. First, it should be emphasized, that S's "doxastic/normative set" of beliefs does not simply comprise all beliefs S has or could have: the doxastic set comprises only beliefs S actually has, the normative set comprises only beliefs which S should have given her evidence (regardless of whether S actually has those beliefs). Since S's normative set depends on the evidence available to S, and since S's doxastic set depends on S's actual beliefs, there might well be beliefs that S could have, that are not contained in S's doxastic/normative set. That means it is entirely possible that there are beliefs, that S could have, but that S neither actually holds, nor are they beliefs that S should hold, given the evidence available to her.

Secondly, a possible criticism of NO DEFEATER is that it leads to scepticism. I imagine, that the argument here goes as follows: for any belief that p, there is always some belief in the doxastic/normative set of S that makes it the case that S should

[137] A nasty problem that looms here is known as the "generality-problem", which arises from the difficulty of providing a principled account of what the "salience" of a process-type boils down to (cf. e.g., Conee & Feldman 1998).

withhold or disbelieve that p*. Perhaps this argument gains some initial plausibility from reflecting on the "messiness" of the doxastic sets of actual believers. But I think we have good reason to resist this line of argument. Notice that this argument would also call into doubt the existence of doxastic and normative defeaters more generally. So, this argument does not apply to NO DEFEATER specifically.

Moreover, it is actually a dubious claim that for any or even for most beliefs of mine, that I take to be justified, we can find elements in my doxastic or normative set, that generate an undercutter in the above sense. It is far from obvious that for many ordinary beliefs of mine, like e.g., my belief that there's a PC in front of me, we can find a belief of mine or a belief I should have, that makes it the case that I should withhold or disbelieve that my belief that there's a PC in front of me was reliably formed.

A major advantage of a condition like NO DEFEATER, compared to other higher-order requirements, is that this condition does not imply that for any of S's beliefs to count as justified, S must hold the respective higher-order belief that that belief was formed reliably (cf. Moon 2018: pp. 254ff.). NO DEFEATER is perfectly compatible with the possibility that S's belief that p is justified, and S e.g., has never even considered whether p*. NO DEFEATER is therefore perfectly compatible with externalist accounts of justified belief, on which it is not required for S's belief that p to be justified that S is or could easily be aware of information that shows that her belief that p was formed in a trustworthy manner (cf. also appendix A).

6.6 APPENDIX F: COGNITION DEFEAT

Here I want to provide some further intuitive motivation for COGNITION DEFEAT. Let's first consider an ordinary case. Intuitively, in normal circumstances, Carla's visual perception of a red block in front of her provides her with justification for the belief in the proposition that there is a red block in front of her. Now it seems plausible to assume that if things are here as they normally are, then it is not the case that Carla should withhold belief that explanatory history of her visual perception

does not involve facts about middle-sized objects and their qualities.[138] Indeed, in this case we might even look for an evolutionary account to support the claim that the explanatory history of visual perception involves facts about middle-sized objects and their qualities. Griffiths and Wilkins (2013: pp. 137-139) have argued that our evolved cognitive mechanisms (like e.g., our visual perception) which we use to form everyday beliefs (like e.g., beliefs about middle sized objects and their qualities) are probably adaptations. And given that they are adaptations, it is hard to see "what…[their] basic evolutionary function could be other than tracking truth" (ibid.). If this is right, then it seems we can be optimistic about the claim that evolutionary explanations concerning our cognitive mechanisms like perception will not make it the case that we should withhold belief that the explanatory history of these mechanisms involved the relevant truths or facts. Therefore, COGNITION DEFEAT is compatible with being justified in believing in the reliability of one's relevant cognitive mechanism in ordinary cases, where we would intuitively judge that these beliefs are indeed justified.

Now, let's return to the scenario featuring Johnny and his visual perception of a red car, familiar from section (1.4.1). At the outset, Johnny is initially justified to believe that the faculty that he uses to arrive at his belief about the object in the driveway (and which he assumes to be ordinary visual perception) is reliable when it comes to producing accurate beliefs about objects like cars in the appropriate conditions. Say e.g., that Johnny then receives convincing information that he has been the subject in an experiment. He receives information that the scientists conducting this experiment have implanted in him a belief-forming faculty that in various circumstances issues beliefs about certain objects in his vicinity and their qualities. However, it seems that the scientists did not design this faculty to produce accurate beliefs about objects like cars in the appropriate, ordinary conditions. Rather, they randomly chose the propositions that this faculty would prompt him to believe in certain circumstances from a list of some true and some false propositions about objects in his vicinity and their qualities.[139] The information he receives also clearly and convincingly tells him

---

[138] I assume here that the facts about red blocks fall under facts about middle-sized objects and their qualities. And a bit further down below, I assume that facts about red cars also fall under facts about middle-sized objects and their qualities too.

[139] This example is a variation of a case presented by Locke (2014: p. 229).

that he has used this faculty in forming the belief about the car in the driveway – and it is also the case that he used this faculty to form the belief.

Johnny has come to believe all of this about himself (or in any case, Johnny should come to believe all of this about himself given his evidence). After coming to believe all of this, it seems that Johnny loses his initial justification for believing that he has used a reliable way of arriving at beliefs about the object in front of him. Why? Because Johnny will have stopped believing (or should have stopped believing) that the facts about objects like blocks and cars are part of the explanatory history of the implanted faculty. This seems like a plausible diagnosis of what makes it the case that Johnny loses his initial justification for believing in the reliability of the relevant cognitive mechanism. This provides COGNITION DEFEAT with some intuitive plausibility.

6.7 APPENDIX G: BEST EXPLANATION

Here I would like to discuss the following question in a bit more detail: When should S withhold belief that the explanatory history of M involves X-facts?[140] Here are two possible answers, which seem to suggest themselves:

> ***NO EXPLANATION***. If there is no explanation available to S of her belief-forming process M that involves (=presupposes, posits, implies or makes likely) the existence of X-facts, then S should withhold belief about the claim that the explanatory history of M involves X-facts.
>
> ***BEST EXPLANATION***. If the best explanation available to S of her belief-forming process M nowhere involves (=presupposes, posits, implies or makes likely) the existence of X-facts, then S should withhold belief about the claim that the explanatory history of M involves X-facts.

NO EXPLANATION says that the lack of an explanation suffices to make it the case that S should withhold belief about the claim that the explanatory history of M involves X-facts. BEST EXPLANATION says that the availability[141] of a certain best

---

[140] This is a question that Locke does not discuss in any detail.

[141] As McCain comments, there are a myriad of proposals for what the "availability" of an explanation amounts to (2016: p. 10). But it is clear that the notion of "availability" involved here needs to be a modest one to be plausible. Let us assume the following working

explanation suffices to make it the case that S should withhold belief about the claim that the explanatory history of M involves X-facts. What makes an explanation "best"? For an explanation to be the best available all that is required is that there is no equally good or better rival explanation available.[142]

To get clearer about what that means, let's look at how this applies to EDAs. In the context of EDAs, what we want to explain is why moral believers have the moral belief-forming faculty they do have and that they use to arrive at their moral beliefs. Now we have an explanation on the table that posits, implies, presupposes or makes likely the proposition that there are non-naturalistically construed moral truths or facts. This is a vindicating moral genealogy. But we also have an explanation on the table that does not involve (=posit, imply, presuppose or make likely) this proposition. This is the debunking moral genealogy. Two explanations are rivals with respect to a proposition if and only if (i) both explanations concern the same explanandum, but (ii) one explanation does not involve a certain proposition, while the other does, and (iii) it is also the case that not both explanations could be true. Therefore, the vindicating and the debunking moral genealogy are rivals.[143] From the

---

assumption of what the "availability" of an explanation comes down to, taken from McCain (2013: p. 303; 2016: p. 10).

Assume that what we want to explain is why moral believers have the moral belief-forming faculty they do have. Let us say that the best explanation available to us involves the proposition that there are non-naturalistically construed moral truths or facts. This part of the explanation and the explanation itself are available to a moral believer if she has the concepts required to understand the proposition and if she is disposed to have a seeming that the proposition is part of the best answer to the question of why she forms moral beliefs in the way she does. According to one popular account, "seemings" are not themselves beliefs or inclinations to believe, but rather *sui generis* mental states with a propositional content and a distinct phenomenology (ibid.; cf. also Tucker 2013). To have the concepts required to understand a proposition and to see it as part of the best explanation one has, all that is required is that the believer has "the ability to understand the question 'why do I have this?' (where 'this' refers demonstratively to her evidence) and the ability to understand when something is the best answer to this question" (McCain 2013: p. 303).

[142] This point too is taken from McCain (2013; 2014; 2016).

[143] It is important to see that the standard responses are not vindicating explanations in this sense, although they do feature some sort of evolutionary explanation. The vindicating explanation we are looking for here is an evolutionary explanation of the development of our moral faculty that features non-naturalistically construed moral facts (and these facts are not superfluous to the explanation). It is an evolutionary explanation of why moral believers have the moral-belief forming mechanism they do have. The standard responses do not feature explanations of why moral believers have developed the moral faculty they do have. Rather, these accounts feature a story of how this faculty would come to be reliable (given that our moral beliefs are mostly true), even though the explanation of the development of

vantage point of the debunker, what matters than for satisfying BEST EXPLANATION is that there is no equally good or better explanation for why moral believers have the moral belief-forming faculty they do have that features the proposition that there are non-naturalistically construed moral truths or facts. Whether one explanation is better than another depends partly on which explanation is better supported by the available evidence, and on which one is more explanatorily virtuous, i.e. which explanation is more parsimonious, has more explanatory power etc.[144]

BEST EXPLANATION also requires one further restriction. We can imagine cases where the best explanation available to S is still a very poor one. For BEST EXPLANATION to be plausible, the presence of an explanation with certain qualities in one case needs to make a tangible epistemic difference to the situation. It does not seem clear that the presence of an exceedingly poor explanation is sufficient to make that difference. Therefore, the kind of explanation we are looking for in BEST EXPLANATION needs to be both (i) sufficiently good, and (ii) it must be the case that there is no equally good or better rival explanation available.

Now we need to settle on either NO EXPLANATION or BEST EXPLANATION. In one sense, NO EXPLANATION is a stronger demand than BEST EXPLANATION, as the former declares that the absence of an explanation is enough for it to be the case that S should withhold while the latter demands this only if an explanation with certain qualities is available to S. On the other hand, there is also a sense in which NO EXPLANATION is a weaker demand than BEST EXPLANATION. To neutralize the defeating potential of NO EXPLANATION it suffices if S comes into the possession of any appropriate explanation. However, to neutralize the defeating potential of BEST EXPLANATION, S needs to come into the possession of a better vindicating explanation.

---

this faculty does not feature non-naturalistically construed moral facts. In other words: a vindicating explanation in the above sense explains why we have a reliable faculty for forming moral beliefs (rather than an unreliable moral faculty or no faculty at all) (Morton 2018: p. 4). And the standard responses do not explain why we have some reliable moral faculty, but rather propose that evolution has imbued us with tendencies to form certain beliefs, and (as it happens) those beliefs are true (ibid.).

[144] And clearly, these two factors are not unrelated.

Now, this already points to a weakness of NO EXPLANATION. NO EXPLANATION declares that the following two cases have different epistemic consequences:

(a). There is no explanation available to S of M that involves X-facts.

(b). S has an explanation of M that involves X-facts, but it is a very poor explanation and S is aware of this or should be aware of this.

NO EXPLANATION entails that S should withhold belief about the claim that the explanatory history of M involved X-facts in (a), while S in (b) (as far as NO EXPLANATION is concerned) is free to believe this claim. However, that is counterintuitive: if the relevant explanation is sufficiently poor, and if S is aware of this (or should be aware of this), then it seems the presence of such a poor explanation should not make a positive epistemic difference. In all cases that fit this general description, if it were true that S should withhold in (a), then S should withhold in (b) too.[145]

To see this, consider the following pair of cases:

(c). There is no explanation available to Johnny concerning his belief-forming mechanism that involves facts about middle-sized objects. Johnny is aware of this or should be aware of this.

(d). There is an explanation available to Johnny concerning his belief-forming mechanism that involves facts about middle-sized objects, but it is a very poor explanation. Johnny is aware of this or should be aware of this.

When we consider (c) and (d), I think it becomes plain that if it were true that Johnny should withhold belief in (c), then he should also withhold belief in (d).[146] However, this result is not in line with NO EXPLANATION. This provides us with a reason to reject NO EXPLANATION.[147]

---

[145] Just to be clear: Since I argue that the debunker should opt for BEST EXPLANATION here, I do not argue that S should withhold belief in either (a) or (b). I am just trying to make explicit what I take to be a counter-intuitive consequence of NO EXPLANATION.

[146] Cases like these also motivate the above restriction of BEST EXPLANATION to explanations that are good *enough*.

[147] I do not regard this point as decisive: perhaps NO EXPLANATION could be amended to deal with cases like these, e.g. by introducing restrictions on the quality of the explanation that S needs to have available.

Nevertheless, let us also consider the choice between NO EXPLANATION and BEST EXPLANATION in the context of EDAs. Here it is rather clear that if you want to set up your evolutionary moral genealogy as an explanatory challenge, then you should opt for BEST EXPLANATION. The reason for this is simple: EDAs are built on the idea that the presence of empirical evidence on the evolution of human morality generates an undercutting defeater. But this empirical evidence plays no role in an explanatory challenge that proceeds via NO EXPLANATION.

I have written above that EDAs are moral genealogies, turning specifically on the evolutionary origin of and influences on our moral beliefs. Every EDA deserving of its name will thus feature a non-superfluous empirical part in it, drawing on findings in recent scientific work on the evolution of human morality. When I say "non-superfluous", what I mean to say is that whether or not your debunking argument is sound should depend partly on whether this empirical story is true. If we opt for NO EXPLANATION, we construct a challenge that does not feature a non-superfluous empirical part at all. Since NO EXPLANATION only requires the absence of a relevant vindicating explanation, the presence of a debunking explanation drawing on what we know or have good reason to believe about the evolution of morality is superfluous to a challenge proceeding via this principle.[148]

BEST EXPLANATION on the other hand, gives evolutionary considerations a real job to do: what the best explanation of our moral belief-forming faculty is, will to a considerable extent depend on which explanation is best supported by the empirical evidence, which will include evidence concerning the evolution of morality.[149]

This point is highly relevant to the present debate of evolutionary debunking: several critics[150] of EDAs have alleged that the empirical part of the arguments is actually redundant, as evolutionary objections collapse into non-empirical epistemological objections to NNMR, which makes the emphasis on the evolutionary influence on

---

[148] That means the presence of a debunking explanation drawing on what we know or have good reason to believe about the evolution of morality is not necessary to mount an explanatory challenge built on NO EXPLANATION. And I take it that part of what makes an argument an EDA is that it *necessarily* features an empirical component of this sort.

[149] This last point of course is reflected in and carried to its extreme by the idealizing assumptions in appendix C, and by assumptions (I-II) in section (1.4.5).

[150] Cf. e.g., Clarke-Doane 2012; Jonas 2016; Klenk 2016; Vavova 2014; White 2010.

morality in the presentation of EDAs grossly misleading. These critics have argued that at their heart, EDAs build on the claim that NNMR cannot explain how our moral beliefs ever non-accidentally land on the moral truth. This is said to be due to the fact that moral truths never cause our moral beliefs and our moral beliefs do not cause or constitute the moral truths. But since this fact is entailed by the metaphysical commitment of moral non-naturalism to the proposition that the moral truths are stance-independent and causally inert, the objection does not depend on the empirical findings of recent scientific work.

The fact that is supposed to make epistemic trouble for the non-naturalist if we opt for NO EXPLANATION is that the moral believer lacks an account of how moral facts figure in the explanation of her moral faculty. However, you might think that this fact arises simply from reflection on the metaphysical commitments of non-naturalism (cf. Crow 2016: p. 380). Hence opting for NO EXPLANATION does not seem to be a viable option for the *evolutionary* debunker.

EDAs hold that the relevant fact that make epistemic trouble for NNMR is crucially connected to the facts about the evolutionary origin of our moral beliefs. In particular, (most) EDAs claim that it is the way that natural selection has causally affected our moral beliefs that creates an epistemic problem. Since the nature of this impact is an empirical matter, the evolutionary objection includes an empirical hypothesis that specifies the nature of this evolutionary influence on our moral beliefs. Therefore, an EDA qua being an EDA must feature a (non-superfluous) empirical part. Since BEST EXPLANATION makes empirical considerations a non-superfluous part of the argument, and since BEST EXPLANATION seems independently more plausible than NO EXPLANATION, I opt for BEST EXPLANATION over NO EXPLANATION in constructing the generic EDA.

Before we proceed, we should also briefly stop to think about whether BEST EXPLANATION is independently plausible. I think it is. If the best explanation available to Johnny concerning his belief-forming mechanism nowhere presupposes, posits, implies or makes likely the existence of facts about middle-sized objects, then Johnny should withhold belief about the claim that the explanatory history of this belief-forming mechanism involves facts about middle-sized objects. This strikes me as highly plausible.

6.8 APPENDIX H: INDEPENDENCE CONSTRAINT

To motivate the INDEPENDENCE CONSTRAINT, consider the following case, again taken from Locke (2014: p. 229), although I have amended it a bit:

> ***CAMMIE'S SPORTS BELIEFS***. Cammie has lots of beliefs about various sports. However, she did not arrive at those beliefs in any of the usual ways (watching television, reading books, browsing the internet, asking friends etc.). Rather, a mad scientist designed and implemented in her a belief-forming faculty (call this her "sports-faculty") that in various circumstances issues beliefs about sports. Moreover, when the scientist was putting this faculty in place, he randomly chose the propositions that this faculty would prompt her to believe in certain circumstances from a list of some true and some false propositions about sports. Cammie has come to justifiably believe all of this about herself. However, Cammie's epistemic guardian angel now comes to the rescue: the angel provides Cammie with a divine sports almanac, that contains all truths about sports. By reading the sports almanac, Cammie has discovered that all the sports propositions that the scientist randomly chose for her to believe happen to be true. Her epistemic guardian angel, whom Cammie justifiably believes to be well meaning, incapable of lying or deceiving and omniscient, assures her that her sports-faculty will continue to produce exclusively true beliefs about sports.

In this case, Cammie can see that the facts that explain her sports faculty were such that they shaped her faculty to be reliable even though the explanatory history of her faculty does not involve facts about sports (cf. ibid.). This is the case, since Cammie can see that once the scientists happened to select only true propositions, and once the guardian angel gives Cammie her assurance, it was guaranteed that Cammie would come to have a faculty that would reliably produce accurate beliefs about sports (cf. ibid.). In that case, it seems that Cammie remains justified in relying on her sports faculty, despite the fact that she believes that the explanatory history of this faculty involves no sports facts (ibid.) This is the case, because Cammie has means independent from her sports faculty to assure herself of the reliability of this faculty (via the guardian angel, the divine sports almanac and via the cognitive faculties that

she uses to take in and store the information contained in the almanac and provided by the angel's testimony).

The INDEPENDENCE CONSTRAINT can also be further motivated in the following way. Let's say there is no epistemic guardian angel and no sports almanac. Let's stipulate that Cammie has no way of assuring herself of the reliability of her sports faculty that is independent of this same faculty. In that case, and if Cammie should still believe that the explanatory history of her sports faculty does not involve sports facts, it seems impermissible for Cammie to believe in the reliability of her sports faculty.

The demand for independent confirmation posited by the INDEPENDENCE CONSTRAINT does not appear to be unreasonably demanding since it only arises in cases where we already have some reason to doubt the reliability of the belief-forming mechanism in question. It does not touch the justification of the ordinary believer who is not confronted with a potentially defeating explanation of M. It only arises in cases where we should withhold belief in the claim that the explanatory history of M involves X-facts. Moreover, it appears that this demand is not unreasonably demanding for the further reason that when it arises, it can at least sometimes be met. Cammie can use other cognitive mechanisms of hers to find out whether her sports faculty produces accurate beliefs about sports.

And the same thing seems true in more ordinary cases. Let's imagine that after running through some tests, Larissa is told by her optometrist that her colour vision might be deceiving her. The tests indicate that Larissa is red-green colour blind. Since people with red-green colour blindness struggle with discriminating red and green hues, after receiving this information, it is plausible that Larissa is no longer permitted to think of herself as reliable in e.g., judging (on the basis of her vision alone, and from a certain distance) that there are no wild strawberries growing on that meadow. What's more, she cannot regain this permission by reasoning:

> But I see no strawberries on that meadow. So, there are no strawberries on that meadow. So, my colour vision is reliable.

This is a clearly fishy response. On the other hand, if Larissa can repeatedly and independently confirm that her colour vision gives her the right results through e.g.,

testimony, then she might be able to regain this permission after all. Therefore, I conclude that the INDEPENDENCE CONSTRAINT looks like a plausible demand.

6.9 Appendix I: The Epistemological Underpinning of the Generic EDA Revisited

Here's a succinct, abstract statement of the epistemological story, which undergirds the generic EDA. For a subject S, a belief-forming mechanism M that forms belief involving the concept X, a belief that p formed via M involving the concept X, and with "X-facts" referring to the facts involving the object/property/relation/kind/etc. picked out by X:

(i) The best explanation available to S for her belief-forming mechanism M nowhere involves (=presupposes, posits, implies or makes likely) the existence of X-facts.

(ii) If the best explanation available to S nowhere involves the existence of X-facts, then S should withhold belief about the claim that the explanatory history of M involves X-facts.

(iii) Therefore, S should withhold belief about the claim that the explanatory history of M involves X-facts.

(iv) If S should withhold belief about the claim that the explanatory history of M involves X-facts, then it is epistemically permissible for S to believe that M is reliable only if her belief is based on a source for justification independent of M.

(v) It is epistemically permissible for S to believe that M is reliable only if her belief is based on a source for justification independent of M.

(vi) S's belief that M is reliable is not based on a source for justification independent of M.

(vii) It is not epistemically permissible for S to believe that M is reliable.

(viii) If it is not epistemically permissible for S to believe that M is reliable, S should withhold or disbelieve that M is reliable.

(ix) S should withhold or disbelieve that M is reliable.

(x) If S is justified in believing that p via M then it is not the case that S should withhold or disbelieve that M is reliable.

(xi) S is unjustified in believing that p via M.

Is it a plausible story? In constructing this story, I have tried to lend some motivation to every step we have taken along our way. I hope this suffices to lend this account some independent plausibility. Nevertheless, we can probe a bit deeper by asking two further questions:

(i) What about the compatibility of this story with contemporary epistemology?

(ii) What distinguishes the challenge generated by this story from global sceptical challenges?

Let's tackle (i) first. Even if the above story is initially plausible, it would clearly not bear well for the story's overall plausibility if it were not compatible with at least significant parts of contemporary epistemology. Let's see how things stand with regards to this.

As I have noted, NO DEFEATER is widely endorsed, and what's more, epistemologists who otherwise disagree a lot with each other endorse it (or versions of it) (cf. Moon 2018: p. 270). The collection of principles BEST EXPLANATION, COGNITION DEFEAT and the INDEPENDENCE CONSTRAINT taken together do not strike me as posing a demand that is not or could not be accommodated by at least many respectable epistemological theories. All that is needed to accommodate this collection of principles is to agree with the following: the availability of a certain kind of explanation makes a difference to whether a subject can continue to believe in the reliability of a cognitive mechanism of hers without having to confirm its reliability independently.

Versions of externalism and internalism about epistemic justification might very well have the room to accommodate this thought. In the next few paragraphs, I will try to give some reason to think that the previous sentence is true. Since it goes beyond the scope of this appendix (or even this thesis) to go through every, most or even many versions of externalism and internalism, I will instead proceed in a rather eclectic manner. I will briefly sketch out why the availability of a certain explanation might well make an epistemically relevant difference on one popular version of each of externalism and internalism respectively.

As I have said in appendix A, I take it that internalists about epistemic justification assert that the potential knowability through reflection of at least some essential part of the justifier for S's belief that p is necessary for S's belief that p to be justified. Externalists deny that this is necessary for epistemic justification.

Now, externalists about epistemic justification, who accept a version of NO DEFEATER, usually build "no-defeater"-conditions into their accounts of epistemic justification. Take e.g., Michael Bergmann's (2006: p. 135) account:

> ***BERGMANN'S PROPER FUNCTIONALISM***. S's belief that p is justified if and only if (i) S does not take the belief that p to be defeated and (ii) the cognitive faculties producing the belief are (a) functioning properly, (b) truth-aimed and (c) reliable in the environments for which they were "designed".[151]

What's important for us here is that Bergmann's account seems compatible with BEST EXPLANATION, COGNITION DEFEAT and the INDEPENDENCE CONSTRAINT. If S receives information that the best explanation of M available to S does not involve the relevant facts, this plausibly provides her with a *prima facie* reason to think that the cognitive faculty M that produces the relevant belief is not truth-aimed. Now assume that M is despite this *in fact* a properly functioning, truth-aimed faculty, that works reliably in the right environment. So, the information that

---

[151] As Mogensen (2014: p. 8) points out, within the context of the debate on evolutionary debunking, there also looms another possibility for utilizing Bergmann's account. According to Bergmann's account, a belief is justified only if it is the product of a cognitive faculty whose function it is to produce true beliefs in the appropriate environment. On one prominent naturalistic account of function, to which Bergmann points to, "the function of a trait is to produce whatever effect has led to selection for that trait within the organism's [i.e., the organism that is the bearer of that trait] evolutionary history" (ibid.; cf. e.g., Milikan 1984 & 1989). Assuming "proper functionalism", and this "selected effect"-account of function, one might then argue that "our moral beliefs are not justified simply because human moral psychology has not been shaped by selection for accuracy" (Mogensen 2014: p.8).
Framed this way, this seems to be an argument for the claim that evolutionary science shows that our moral beliefs are unjustified and never were justified to begin with. And this is quite different from a challenge framed in terms of epistemic defeat, where it is assumed that our beliefs were initially justified, and where incoming information is said to affect our initial justification via changing what we believe or should believe (cf. ibid.).
However, we might reframe this challenge in terms of defeat. First, our reasoning could start with e.g. an appeal to modesty: even if the empirical hypothesis involved in our EDA is very well supported, we should perhaps still not fully commit ourselves to the claim that evolutionary science simply shows our moral beliefs to never have been justified. Instead we could adopt the view that given the currently best explanation of our moral faculty, we should take it that we are no longer entitled to our moral beliefs.

makes it the case that the best explanation of M available to S does not involve the relevant facts is *misleading* in this respect. Nonetheless, given everything that S believes or should believe, it seems that receiving this information then plausibly furnishes S with a doxastic or normative undercutting defeater for her belief that p produced via M, at least in the absence of independent means for S to confirm the reliability of M. Since it seems that evidence of unreliability can be defeating, even if it is misleading, it also seems that a satisfying externalist account of justification should have room for accommodating this thought.[152]

The "no defeater"- condition that Bergmann builds into his account, and his commitment to a picture of defeat, which is basically in line with NO DEFEATER, should enable his brand of Proper Functionalism to accommodate this thought. The INDEPENDENCE CONSTRAINT then formulates a plausible constraint on what it takes for S to be in a position to reinstate her justification. S needs to have a way of independently confirming that M is after all reliable. If S can confirm this, then S has reason to think that M is truth-aimed after all. And this neutralizes the defeating force of the best explanation for M via giving S reason to think that M is truth-aimed.

I thus conclude that the above triplet of principles seems compatible with Bergmann's account of epistemic justification. On this account, the availability of a certain kind of explanation makes the appropriate epistemic difference via calling into doubt whether S's cognitive faculty is truth-aimed, and by so (potentially) generating an undercutting defeater for S's belief that p.

Now, this result seems to have a general applicability beyond Bergmann's account. If S finds out that the best explanation of M does not involve the relevant facts, this plausibly provides her with a *prima facie* reason to think that the cognitive faculty M that produces the relevant belief is not truth-aimed. Moreover, it's eminently plausible that having a *prima facie* reason to think that M is not aimed at the truth in producing beliefs should be of epistemic consequence for S. This result (= receiving information that gives S a reason to think that M is not truth-aimed is of epistemic consequence) is likely to be in line with versions of externalism beyond Bergmann's. Furthermore, it is also a result that seems to be in line with many versions of internalism.

---

[152] Cf. Goldman (1979).

Take for instance Michael Huemer's Phenomenal Conservativism, which very roughly holds:

> ***PHENOMENAL CONSERVATIVISM***. If it seems to S that p, then, in the absence of defeaters, S thereby has (at least some) justification for believing that p. (Huemer 2007: p. 30; cf. also Huemer 2006: pp. 148-149; Huemer 2005: Ch. 5 & 2014).[153]

What matters for our purposes here is that for a seeming to justify S in believing that p, it must not be the case that given her background beliefs, S believes that there's no appropriate relationship between the fact that p and S's seeming that p (cf. McCain 2016: p. 12). Let's say that the best explanation available to S for why S has the seeming that p nowhere involves the fact that p. Again, it seems plausible to say that receiving this information is apt to give S a reason to think that there is no appropriate relationship between the fact that p and S's seeming that p. Let's say there is no better explanation for why it seems to S that p that does involve the fact that p. Absent other means to show that S's seeming that p indeed does stand in an appropriate relationship with the fact that p, it very much appears that S loses her justification to

---

[153] "Seeming" states, understood in a broad sense, include perceptual, intellectual, memory, and introspective appearances (Huemer 2014: section 1a.). According to Phenomenal Conservativism, justification is grounded in the way things appear or seem to the subject (McCain 2016: p. 2). As Huemer writes:

"[Phenomenal Conservativism]… fits with an internalistic form of foundationalism—that is, the view that some beliefs are justified non-inferentially (not on the basis of other beliefs), and that the justification or lack of justification for a belief depends entirely upon the believer's internal mental states. The intuitive idea is that it makes sense to assume that things are the way they seem, unless and until one has reasons for doubting this." (Huemer 2014: introduction).

"Seemings" are not themselves beliefs or inclinations to believe, but rather *sui generis* mental states with a propositional content and a distinct phenomenology, that can confer (non-inferential) justification upon beliefs, and there are a number of competing accounts of seeming states (cf. Tucker 2013).

Huemer's version of moral intuitionism is also tied up with his version of Phenomenal Conservativism. As Huemer uses the term, "intuitions" are intellectual seemings, that can (in the absence of defeaters) confer non-inferential justification on beliefs. Although moral intuitions on this account are epistemically significant, in that they can provide non-inferential justification for our moral beliefs (in the absence of defeaters), Huemer's version of intuitionism eschews any allegiance to some of the more controversial commitments of more traditional versions of moral intuitionism, such as "claims to direct insight, indefeasibility, certainty and infallibility" (Hermann 2017: p. 830). Note also that Huemer's moral intuitionism contrasts with other recent versions of this view, which hold that "intuitions" are (roughly) non-inferentially justified beliefs (Audi 2004; Shafer-Landau 2008).

believe that p because it seems to her that p. So, with a few minor alterations, it seems that our triplet of principles could be accommodated by a major version of epistemic internalism as well.[154]

Sketchy as these remarks may be, I hope they nonetheless achieve the purpose of providing you with some reason to think that the above story is an account worth taking epistemologically seriously. The principles invoked in this story seem to be independently plausible and compatible with a couple of the major epistemological theories on epistemic justification.

Now, let's tackle (ii): in the context of both the debate about evolutionary debunking (cf. Berker 2014; Vavova 2014 & 2015) and in epistemology more generally, epistemic principles are often criticized for entailing global scepticism. This is one of the most popular ways for casting suspicion on any given epistemic principle.

Does this charge have any bite against the collection of principles involved in our epistemological story, which taken together formulate an explanatory constraint on

---

[154] Since the epistemological story presented here rests on claims about the epistemic relevance of explanations, and since a recent form of internalism puts explanations centre stage, you may wonder about the connection between the two. Kevin McCain's (2013; 2014; 2016) *explanationist evidentialism* very roughly holds that your belief is justified just in case your belief "fits" the evidence, where a belief, p, fits the evidence, e, at time t, just in case "p is part of the best explanation available to S at t for why S has e" (McCain 2014: p. 63). I have obviously drawn a lot on McCain's exposition and defence of the view, which is attested by the many times I have quoted him while developing the story above.
Now, the explanatory constraints featured in the epistemological story above are certainly a lot weaker and more indirect than the claims that explanationist evidentialism makes about the connection between justification and explanation. Explanationist evidentialism defines justification in terms of the best available explanation. All that is needed for our above account is the following: the availability of a certain kind of explanation makes a difference to whether a subject can continue to believe in the reliability of a cognitive mechanism of hers without having to confirm its reliability independently. To think that the presence of a certain kind of explanation is one of the things that can give rise to a need for independent confirmation of your reliability on a subject matter, you do not need to endorse explanationist evidentialism.
On the other hand, if explanationist evidentialism is true, then you can develop a way more straightforward version of the story above. Here's a rough sketch of this version:

(i). S is justified to believe that p at time t if and only if p is part of the best explanation available to S at t for why S has her total evidence e.
(ii). p is not part of the best explanation available to S at t for why S has e.
(iii). Therefore, S is not justified to believe that p at t.

This version of the story resembles the core idea behind Lutz's EDA (2017).

which claims we can believe with justification (i.e., BEST EXPLANATION; COGNITION DEFEAT; INDEPENDENCE CONSTRAINT)?

I have already noted above that the account we are considering is distinct from global sceptical challenges in allowing for a default entitlement to the belief that our cognitive mechanisms are reliable. Global sceptical challenges usually invoke constraints on justification that aim to establish the point that we never were or are justified to believe any claim at all.

Furthermore, in contrast to sceptical challenges, where it is hard to see how we could ever be in a position to exclude the scenarios, which motivate these doubts (e.g., we cannot know that we are not radically deceived), there are straightforward ways in which challenges built on the above story can be met:

(a). Turn the debunking explanation into a vindicating explanation. That is to show how X-facts are after all involved in the supposedly debunking explanation of M.

(b). Present an equally good or better vindicating explanation. That is to show that there is an equally good or better explanation available to S for M that does involve X-facts.

(c). Show that although the explanatory history of M does not involve X-facts, this is no reason for concern, as S has M-independent means of showing that M is reliable.

If you can do any of (a)-(c), then as far as the above story is concerned, no (undefeated) defeater for S's belief that p has been generated. As has been noted in the literature on the epistemology of scepticism many times, it is characteristic of global sceptical challenges that we lack any clear idea of how we could ever be in a position to exclude the possibility that we are in a scenario that gives rise to these kinds of all-encompassing doubts (cf. Klein 2015: section 1). How can you e.g., show that you are not deceived about everything by a Cartesian demon, or that you are not a brain-in-a-vat?[155]

[155] Of course, some philosophers have tried to argue that you can show that you are not in a sceptical scenario. Moore (1939) is the classic example here. But then again, following Wittgenstein (1969) a substantial part of the literature on scepticism has been dedicated to showing what is wrong with Moore's response.

On a related note, one could also argue that the above story does not lend any support to a globally sceptical position since the sceptic's case for the claim that the best explanation available to us for all our belief-forming mechanisms does not involve the relevant facts has not been sufficiently supported by anyone.[156]

A further point worth emphasizing is that while empirical facts and claims can make a difference to what the best explanation for a belief-forming faculty of ours is, this point does not seem to have a place in the argument of the global sceptic. It is not the case that the global sceptic has good empirical evidence that we are deceived on a large scale and that this gives rise to sceptical doubts. It is rather the claim that we cannot exclude the possibility that we are so deceived that gives rise to these doubts. For example, the sceptic does not present you with experimental data, which suggests that despite what you think you perceive, you may not actually have hands. It follows that if we have reason to think that for a given application of the above epistemological story, the argument that builds on this story features an empirical component that is not superfluous to its soundness, then we have reason to think that this argument is not a globally sceptical one.

Therefore, I conclude that the above story not only has some independent initial plausibility, but also gains further support from the observation that (i) it is compatible with major epistemological theories of epistemic justification on both sides of the externalism/internalism divide, and (ii) it is clearly distinct from globally sceptical challenges.

---

[156] This point is inspired by McCain (2014: p. 131), who advertises explanationist evidentialism for its anti-sceptical capabilities. Here McCain assesses the hypothesis of external world-scepticism in comparison with a common-sense hypothesis and asks which of the two would be the better explanation available to an ordinary person for the total evidence in her possession. He proceeds by evaluating the two hypotheses with respect to a number of explanatory virtues. In the end, he concludes that while it is not clear that the common-sense hypothesis scores higher with respect to every explanatory virtue he proposes, it is nonetheless reasonable to conclude that it is overall better.

7 REFERENCES

Alexander, D. 2017. Unjustified Defeaters. Erkenntnis 82/4: 891–912.

Audi R. 2004. The good in the Right: A Theory of Intuition and Intrinsic Value. Princeton: Princeton University Press.

Ayer, A.J. 1946. Language, Truth, and Logic. 2nd Edition. London: Gollancz.

Ayer, A.J. 1954. On the Analysis of Moral Judgements. In: Ayer, A.J. Philosophical Essays. London: Macmillan: 231-49.

Baker-Hytch, M. Manuscript. On the Evolutionary Epistemological Challenge to Moral Realism.

Baker-Hytch, M. 2014. Religious Diversity and Epistemic Luck. International Journal for Philosophy of Religion 76/2: 171-191.

Baker-Hytch, M. 2017. Epistemic Externalism in the Philosophy of Religion. Philosophy Compass 12/4: 1-12.

Baker-Hytch, M. & Benton, M. 2015. Defeatism Defeated. Philosophical Perspectives 29/1: 40-66.

Barkhausen, M. 2016. Reductionist Moral Realism and the Contingency of Moral Evolution. Ethics 126 (3): 662-689.

Becker, K. 2012. Methods and How to Individuate Them. In: Becker, K. & Black, T. (Eds.) The Sensitivity Principle in Epistemology. Cambridge: Cambridge University Press: 81-98.

Becker, K. & Black, T. (Eds.) 2012. The Sensitivity Principle in Epistemology. Cambridge: Cambridge University Press.

Bedke, M. 2009. Intuitive Non-naturalism Meets Cosmic Coincidence. Pacific Philosophical Quarterly 90/2: 188–209.

Bergmann, M. 1997. Internalism, Externalism and the No-Defeater Condition. Synthese, 110/3: 399-417.

Bergmann, M. 2005. Defeaters and Higher-Level Requirements. The Philosophical Quarterly 55/220: 419-436.

Bergmann, M. 2006. Justification Without Awareness: A Defence of Epistemic Externalism. New York: Oxford University Press.

Berker, S. 2014. Does Evolutionary Psychology show that Normativity is Mind-Dependent? In: http://scholar.harvard.edu/files/sberker/files/berker-evo-psych-mind-dep.pdf?m=1419638931 [Date of Access: 24.02.2017].

Bogardus, T. 2014. Knowledge Under Threat. Philosophy and Phenomenological Research 88/2: 289-313.

Bogardus, T. 2016. Only all Naturalists Should Worry About Only One Evolutionary Debunking Argument. Ethics 126: 636–61.

BonJour, L. 1980. Externalist Theories of Empirical Knowledge. Midwest Studies in Philosophy 5: 135-150.

Braddock, M., Mogensen, A. L. & Sinnott-Armstrong & W. 2012. Online discussion of Clarke-Doane 2012. Online accessible at: {http://peasoup. typepad.com/peasoup/2012/03/ethics-discussions-at-pea-soup-justin-clarke-doanes-morality-andmathematics-the-evolutionary-challe-1.html} [Date of Access: 3rd of April 2017].

Brosnan, K. 2011. Do the evolutionary origins of our moral beliefs undermine moral knowledge? Biology and Philosophy 26/1: 51-64.

Cassam, Q. 2016. Vice Epistemology. The Monist 99/2: 159-180.

Clarke-Doane, J. 2012. Morality and Mathematics: The Evolutionary Challenge. Ethics 122: 313–340.

Clarke-Doane, J. 2015. Justification and Explanation in Mathematics and Morality. In: Shafer-Landau, R. (Ed.) Oxford Studies in Metaethics 10. New York: Oxford University Press: 80-103.

Clarke-Doane, J. 2016. Debunking and Dispensability. In: Sinclair, N. & Leibowitz, U. (Eds.) Explanation in Ethics and Mathematics: Debunking and Indispensability. Oxford: Oxford University Press: 23-36.

Comesaña, J. 2007. Knowledge and Subjunctive Conditionals. Philosophy Compass 2/6: 781-791.

Conee, E. & Feldman, R. 1998. The Generality Problem for Reliabilism. Philosophical Studies 89/1: 1-29.

Conee, E. & Feldman, R. 2004. Evidentialism. Essays in Epistemology. New York: Oxford University Press.

Crow, D. 2016. Causal Impotence and Evolutionary Influence: Epistemological Challenges for Non-Naturalism. Ethical Theory and Moral Practice 19: 379-395.

Cuneo, T. 2011. Reidian Metaethics: Part I & Part II. Philosophy Compass 6 (5):333-349.

Davidson, M. 2013. God and Other Necessary Beings. In: Stanford Encyclopaedia of Philosophy. Online accessible at {https://plato.stanford.edu/entries/god-necessary-being/} [Date of Access: 25.7.2018].

DeRose, K. 1995. Solving the Sceptical Problem. The Philosophical Review 104/1: 1-52.

Dogramaci, S. 2012. Reverse Engineering Epistemic Evaluations. Philosophy and Phenomenological Research 84/3: 513-530.

Dogramaci, S. 2015. Communist Conventions for Deductive Reasoning. Noûs 49/4: 776-799.

Dworkin, R. 1996. Objectivity and Truth: You'd Better Believe it. Philosophy and Public Affairs 25/2: 87–139.

Enoch, D. 2010. The epistemological challenge to meta-normative realism: how best to understand it, and how to cope with it. Philosophical Studies 148: 413–38.

Enoch, D. 2011. Taking Morality Seriously. Oxford: Oxford University Press.

Faraci, D. Manuscript. Knowledge, Necessity and Defeat. Online accessible at: http://davidfaraci.com/ip/defeat.pdf [Date of Access: 31.03. 2018]

Fitzpatrick, W.J. 2015. Debunking evolutionary debunking of ethical realism. Philosophical Studies172 (4): 883-904.

Fitzpatrick, W.J. 2016. Misidentifying the Evolutionary Debunkers' Error: Reply to Mogensen. Analysis 76 (4): 433-437.

Gettier, E. 1963. Is Justified True Belief Knowledge? Analysis 23(6): 121-123.

Gibbard, A. 2011. How Much Realism? Evolved Thinkers and Normative Concepts. In: Shafer-Landau, R. (Ed.) Oxford Studies in Metaethics 6. New York: Oxford University Press: 6-33.

Goldberg, S. 2014. Does Externalist Epistemology Rationalize Religious Commitment? In: Callahan, L.F. & O'Connor, T. (Eds.) Religious Faith and Intellectual Virtue. Oxford: Oxford University Press: 279–298.

Goldman, A. 1979. What Is Justified Belief? In: Pappas, G. (Ed.) Justification and Knowledge. Dordrecht: D. Reidel: 1-23.

Goldman, A. 1986. Epistemology and Cognition. Cambridge, MA: Harvard University Press.

Goldman, A. & Beddor, B. 2015. Reliabilist Epistemology. In: Stanford Encyclopaedia of Philosophy. Online accessible at {https://plato.stanford.edu/entries/reliabilism/} [Date of Access: 25.7.2018].

Griffiths, P. & Wilkins, J. (2013). Evolutionary Debunking Arguments in Three Domains: Fact, Value, and Religion. In: Maclaurin, J. & Dawes, G. (Eds.) A New Science of Religion. Routledge: 132-146.

Hanson, L. 2017. The Real Problem with Evolutionary Debunking Arguments. Philosophical Quarterly 67/268: 508-533.

Hawthorne, J. 2004. Knowledge and Lotteries. Oxford: Oxford University Press.

Hermann, J. 2017. Sinnott-Armstrong's Empirical Challenge to Moral Intuitionism: A Novel Critique. Ethical Theory and Moral Practice 20: 829-842.

Hookway, C. 1994. Cognitive Virtues and Epistemic Evaluations. International Journal of Philosophical Studies 2: 211-227.

Huemer, M. 2005. Ethical Intuitionism. Palgrave Macmillan: Basingstoke.

Huemer, M. 2006. Phenomenal Conservatism and the Internalist Intuition. American Philosophical Quarterly 43/2: 147-158.

Huemer, M. 2007. Compassionate Phenomenal Conservatism. Philosophy and Phenomenological Research 74/1: 30-55.

Huemer, M. 2014. Phenomenal Conservatism. In: Internet Encyclopaedia of Philosophy. Online accessible at {http://www.iep.utm.edu/phen-con/} [Date of access: 2.4.2017].

Kramer, M. 2009. Moral Realism as a Moral Doctrine. Hoboken: Wiley-Blackwell.

Ichikawa, J. 2011. Quantifiers, Knowledge, and Counterfactuals. Philosophy and Phenomenological Research 82/2: 287-313.

Ishikawa, J. & Steup, M. 2017. The Analysis of Knowledge. In: Stanford Encyclopaedia of Philosophy. Online accessible at {https://plato.stanford.edu/entries/knowledge-analysis/} [Date of Access: 25.7.2018].

Isserow, J. 2018. Evolutionary Hypotheses and Moral Scepticism. Erkenntnis: 1-21. https://doi.org/10.1007/s10670-018-9993-8.

Jonas, S. 2016. Access problems and explanatory overkill. Philosophical Studies: 1-12. doi 10.1007/s11098-016-0807-z.

Joyce, R. 2006. The Evolution of Morality. Cambridge, MA: MIT Press.

Joyce, R. 2013. Irrealism and the Genealogy of Morals. Ratio 26: 351-372.

Joyce, R. 2016a. Evolution, Truth-Tracking and Moral Scepticism. In: Joyce, R. Essays in Moral Scepticism. Oxford: Oxford University Press: 142-158.

Joyce, R. 2016b. Reply: Confessions of a Modest Debunker. In: Sinclair, N. & Leibowitz, U. (Eds.) Explanation in Ethics and Mathematics: Debunking and Indispensability. Oxford: Oxford University Press: 124-145.

Joyce, R. 2016c. The Many Moral Nativisms. In: Joyce, R. Essays in Moral Scepticism. Oxford: Oxford University Press: 122-141.

Kahane, G. 2011. Evolutionary Debunking Arguments. Noûs 14/1: 103–25.

Klein, P. 2015. Scepticism. In: Stanford Encyclopaedia of Philosophy. Online accessible at {https://plato.stanford.edu/entries/skepticism/#Rel} [Date of Access: 25.7.2018].

Klenk, M. 2017. Old Wine in New Bottles. Ethical Theory and Moral Practice 20 (4): 781-795.

Lackey, J. 2014. Taking Religious Disagreement Seriously. In: Callahan, L.F. & O'Connor, T. (Eds.) Religious Faith and Intellectual Virtue. Oxford: Oxford University Press: 299-316.

Lasonen-Aarnio, M. 2010. Unreasonable Knowledge. Philosophical Perspectives 24/1: 1-21.

Law, S. 2016. The X-claim Argument Against Religious Belief. Religious Studies: 1-21. doi:10.1017/S0034412516000330.

Lepock, C. 2011. Unifying the Intellectual Virtues. Philosophy and Phenomenological Research 83/1: 106-128.

Lewis, D. 1973. Counterfactuals. Cambridge, MA: Harvard University Press.

Levy, A. & Levy, Y. 2016. The Debunking Challenge to Realism: How Evolution (Ultimately) Matters. Journal of Ethics & Social Philosophy, Discussion Note: 1-7.

Littlejohn, C. 2012. Justification and the Truth-Connection. Cambridge: Cambridge University Press.

Locke, D. 2014. Darwinian Normative Scepticism. In: Bergmann, M & Kain, P. (Eds.): Challenges to Moral and Religious Belief: Disagreement and Evolution. New York: Oxford University Press: 220-236.

Lutz, M. 2017. What Makes Evolution a Defeater? Erkenntnis: 1-22. doi: 10.1007/s10670-017-9931-1.

Millikan, R. 1984. Language Thought and Other Biological Categories. Cambridge, MA: MIT Press.

Millikan, R. 1989. In Defence of Proper Functions. Philosophy of Science 56: 288-302.

McCain, K. 2013. Explanationist Evidentialism. Episteme 10/3: 299-315.

McCain, K. 2014. Evidentialism and Epistemic Justification. Abingdon: Routledge.

McCain, K. 2016. Explanationist Aid for Phenomenal Conservatism. Synthese: 1-16. DOI 10.1007/s11229-016-1064-6.

McPherson, T. 2012. Ethical Non-Naturalism and the Metaphysics of Supervenience. In: Shafer-Landau, R. (Ed.) Oxford Studies in Metaethics 7. New York: Oxford University Press: 205-234.

McPherson, T. 2015. Supervenience in Ethics. In: Stanford Encyclopaedia of Philosophy. Online accessible at {https://plato.stanford.edu/entries/supervenience-ethics/} [Date of Access: 25.7.2018].

Miller, A. 2003. An Introduction to Contemporary Metaethics. London: Polity.

Mogensen, A. L. 2014. Evolutionary Debunking Arguments in Ethics. D. Phil. Thesis: University of Oxford.

Mogensen, A.L. 2015. Evolutionary Debunking Arguments and the Proximate/Ultimate Distinction. Analysis 75: 196–203.

Mogensen, A.L. 2016a. Contingency Anxiety and the Epistemology of Disagreement. Pacific Philosophical Quarterly 97: 590–611.

Mogensen, A.L. 2016b. Do Evolutionary Debunking Arguments Rest on a Mistake About Evolutionary Explanations? Philosophical Studies 173: 1799–1817.

Mogensen, A.L. 2017. Disagreement in Moral Intuitions as Defeaters. The Philosophical Quarterly 67/267: 282-302.

Moon, A. 2016. Debunking Morality: Lessons from the EAAN Literature. Pacific Philosophical Quarterly: 1-19. DOI: 10.1111/papq.12165.

Moon, A. 2018. How to Use Cognitive Faculties You Never Knew You Had. Pacific Philosophical Quarterly 99: 251-275.

Moore, G.E. 1939. Proof of an External World. Proceedings of the British Academy 25: 273-300.

Morton, J. 2018. When Do Replies to the Evolutionary Debunking Argument Against Moral Realism Beg the Question? Australasian Journal of Philosophy: 1-16. DOI:10.1080/00048402.2018.1455718

Nagel, T. 2012. Mind and Cosmos. New York: Oxford University Press.

Nozick, R. 1981. Philosophical Explanations. Cambridge, MA: Harvard University Press.

Pappas, G. 2014. Internalist vs. Externalist Conceptions of Epistemic Justification. In: Stanford Encyclopaedia of Philosophy. Online accessible at {https://plato.stanford.edu/entries/justep-intext/} [Date of Access: 25.7.2018].

Parfit, D. 2011. On What Matters. Volume 2. Oxford: Oxford University Press.

Peels, R. 2017. Responsible Belief and Epistemic Justification. Synthese 194: 2895–2915.

Plantinga, A. 2000. Warranted Christian Belief. New York: Oxford University Press.

Pollock, J.L. 1986. Contemporary Theories of Knowledge. Totowa, NJ: Rowman and Littlefield.

Pritchard, D. 2009. Safety-based Epistemology: Wither Now? Journal of Philosophical Research 34: 33–45.

Pritchard, D. 2012a. Anti-luck Virtue Epistemology. Journal of Philosophy 109/3: 247–279.

Pritchard, D. 2012b. Epistemological Disjunctivism. Oxford: Oxford University Press.

Rabinowitz, D. 2011. The Safety Condition for Knowledge. In: Internet Encyclopaedia of Philosophy. Online accessible at {http://www.iep.utm.edu/safety-c/} [Date of access: 2.4.2017].

Roland, J. & Cogburn, J. 2011. Anti-Luck Epistemologies and Necessary Truths. Philosophia 39: 547–561.

Rosen, G. Manuscript. What is Normative Necessity?

Ruse, M. 1986. Taking Darwin Seriously. Oxford: Blackwell Publishing, 1986.

Russell, B. 1948. Human Knowledge: Its Scope and Limits. London: George Allen and Unwin.

Sainsbury, R. 1997. Easy Possibilities. Philosophy and Phenomenological Research 57/4: 907-919.

Sayre-McCord, G. Moral Realism. In: Stanford Encyclopaedia of Philosophy. Online accessible at {https://plato.stanford.edu/entries/moral-realism/} [Date of Access: 25.7.2018].

Schafer, K. 2010. Evolution and Normative Scepticism. Australasian Journal of Philosophy 88/3: 471-488.

Schechter, J. 2018. Explanatory Challenges in Metaethics. In: McPherson, T. & Plunkett, D. (Eds.): Routledge Handbook of Metaethics. Routledge: 443-459.

Shafer-Landau, R. 2003. Moral Realism: A Defence. New York: Oxford University Press.

Shafer-Landau, R. 2008. Defending Ethical Intuitionism. In: Sinnott-Armstrong, W. (Ed.) Moral Psychology. Volume 2. Cambridge, MA: MIT Press: 83-95.

Sober, E. 1984. The Nature of Selection: Evolutionary Theory in Philosophical Focus. Chicago: University of Chicago Press.

Sosa, E. 1999. How to Defeat Opposition to Moore. Philosophical Perspectives 33/13: 141–153.

Srinivasan, A. 2015. Normativity Without Cartesian Privilege. Philosophical Issues 25: 273-299.

Sripada, C. Nativism and Moral Psychology: Three Models of the Innate Structure That Shapes the Contents of Moral Norms. In: Sinnott-Armstrong, W. (Ed.): Moral Psychology. Volume 1. Cambridge, MA: MIT Press: 319-344.

Street, S. 2006. A Darwinian Dilemma for Realist Theories of Value. Philosophical Studies 127 (1): 109–66.

Street, S. 2008. Reply to Copp: Naturalism, Normativity, and the Varieties of Realism Worth Worrying About. Philosophical Issues 18: 207–228.

Street, S. 2011. Mind-Independence Without the Mystery: Why Quasi-Realists Can't Have It Both Ways. In: Shafer-Landau, R. (Ed.) Oxford Studies in Metaethics 6. New York: Oxford University Press:1-32.

Sudduth, M. 2008. Defeaters in Epistemology. In: Internet Encyclopaedia of Philosophy. Online accessible at {http://www.iep.utm.edu/ep-defea/} [Date of access: 2.4.2017].

Tucker, C. 2011. Phenomenal Conservatism and Evidentialism in Religious Epistemology. In: Clark, K.J. & VanArragon, R.J. (Eds.) Evidence and Religious Belief. Oxford: Oxford University Press: 52-73.

Tucker, C. 2013. Seemings and Justification: An Introduction. In: Tucker, C. (Ed.) Seemings and Justification: New Essays on Dogmatism and Phenomenal Conservatism. New York: Oxford University Press: 1-29.

Väyrynen, P. Forthcoming. The Supervenience Challenge to Non-Naturalism. In: McPherson, T. & Plunkett, D. (Eds.) Routledge Handbook of Metaethics. Routledge.

Vavova, K. 2014. Debunking Evolutionary Debunking. In: Shafer-Landau, R. (Ed.) Oxford Studies in Metaethics 9. New York: Oxford University Press: 76-101

Vavova, K. 2015. Evolutionary Debunking of Moral Realism. Philosophy Compass 10/2: 104–116.

Vogel, J. 2012. The Enduring Trouble with Tracking. In: Becker, K. & Black, T. (Eds.) The Sensitivity Principle in Epistemology. Cambridge: Cambridge University Press: 122-151.

White, R. 2010. You just believe that because…Philosophical Perspectives 24: 573-615.

Wielenberg, E. 2010. On the Evolutionary Debunking of Morality. Ethics 120/3: 441-464.

Wielenberg, E. 2014. Robust Ethics. Oxford: Oxford University Press.

Wielenberg, E. 2016. Ethics and Evolutionary Theory. Analysis Reviews 76/4: 502–515.

Williamson, T. 2000. Knowledge and its Limits. New York: Oxford University Press.

Williamson, T. 2009. Reply to Goldman. In: Greenough, P. & Pritchard, D. (Eds.) Williamson on Knowledge. Oxford: Oxford University Press: 305-312.

Wittgenstein, L. 1969. On Certainty. Harper Torchbooks.