A DEFENCE OF THE EVOLUTIONARY DEBUNKING ARGUMENT

by

MAN HIM IP

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

Department of Philosophy
School of Philosophy, Theology and Religion
College of Arts and Law
University of Birmingham
August 2020

ABSTRACT

In this thesis, I will explore the epistemological evolutionary debunking arguments in meta-ethics (EDA). I will defend these arguments by accomplishing two tasks: (1) I will offer the best (i.e., most detailed and precise) way to understand the EDA and (2) I will also respond to two strongest objections to the EDA.

Firstly, in **Part I** of this thesis, I will offer my account of how the EDA should be best formulated. I will start from how evolution has significantly influenced our moral beliefs. I will then explain why, due to their evolutionary origin, our moral beliefs are not tracking the moral truth reliably. Furthermore, I will argue that the fact that our moral beliefs are not tracking the moral truth also provides an undercutting and a higher-order defeater for those beliefs. As a result, I will conclude that the epistemic status of our moral beliefs is undermined because of these two kinds of defeaters.

Secondly, in **Part II** of this thesis, I will turn my attention to the two strongest objections to the EDA – the Conceptual Truth Objection and the Third Factor Objection. I will first offer two responses to the Conceptual Truth Objection and, based on these two responses, I will argue that the Conceptual Truth Objection fails as a challenge to the EDA. I will then also argue that there currently doesn't exist an acceptable version of the Third Factor Objection and it is also unlikely that such a version could be constructed in the future. I will finally conclude that both objections are problematic and they are thus unable to give us reason to doubt the EDA.

ACKNOWLEDGEMENT

Such a long journey. It is impossible that I could have completed this journey on my own. I would like to express my sincere appreciation to those who have been part of my amazing experience of accomplishing this research project.

The first person that I am deeply indebted to is my lead supervisor Dr. Jussi Suikkanen. I have had a great fortune to learn from Jussi, who has continuously provided me with his insightful comments on philosophy and also writing advice in the last four years. He really is a role model to me in being a philosopher, a supervisor and a teacher. I am thankful beyond words for his patience, kindness and sympathy. I am so blessed to have Jussi as my supervisor.

I am also grateful to my co-supervisor Prof. Scott Sturgeon for providing me with his invaluable comments. Scott's comments contribute to a significant change in the direction of this research project. I sincerely believe that my research project has been much improved since then. In addition, I must also thank my examiners Prof. Alastair Wilson and Prof. Neil Sinclair.

I would like to thank the philosophers who have contributed to my philosophical growth with their thoughts, comments and encouragement since my undergraduate years: Prof. Lei Zhong, Prof. Pekka Väyrynen, Prof. Leo Kam-Ching Cheung, Prof. Kai-Yee Wong, Prof. Yong Huang, Prof. Chong-Fuk Lau, Dr. Kai-Chiu Ng and Prof. Chi-Keung Chan. I owe a special debt to Prof. Tien-Chi Chen, who gave me the chance to work for him

as a teaching assistant. Without him, I would have given up my academic career a long time ago.

Thanks to Chun-Yin “Salt” Yeung, Chun-Man Kwong, Howard Mok, Ko-Hung Kuan and Tien-Chun Lo for their friendly support in philosophy and also in my life. I hope that we can meet together in the future – in Oxford, in Taiwan, in Hong Kong, in Birmingham, or in London.

Thanks to my friends Harris, Zuson, Simon, Marc, Chi-Ho, Kyle, Ho-Hin, Dickson, Alex, Joe Hung, Hong-Ting, Ho, Joanne, Polly, Connie, Tsz-Yan, Panda, Viv, Edmund, Wing, Tsz Mei, Kum, Tony, Hei Man and John for their genuine friendship and support.

Last but not least, I am grateful beyond measure to my parents – Samson and Carrie – for their support. I hope that you would be proud of me. I would like to dedicate my thesis to them.

CONTENT

Chapter 1

Introduction

Even though we deeply care about morality and often take it for granted, at the same time we also find it difficult to understand the historical origins of human morality and the nature of our moral beliefs. In recent decades, many philosophers have attempted to elucidate the latter by relying on the evolutionary theory.[1] As a consequence, these philosophers have suggested that we can, for example, explain our different moral intuitions and moral beliefs by describing their evolutionary origins.

In order to get a sense of how this could be done, let us consider two slightly different moral problems. We can begin with the classic trolley problem.[2] A runaway trolley is heading down a track and it is about to hit five tied-up people lying on the track. All of them will be killed if the trolley continues on its course. The only way to save these people is to pull a lever that will divert the trolley onto a sidetrack. As a consequence of doing so, however, another individual, who is tied-up and lying on the sidetrack, would be killed by the trolley. We may then ask: Are you permitted to pull the lever? When this question is asked, most people think that we are permitted to pull the lever to save five lives even if this means having to sacrifice the one person (Greene 2016: 175; Singer 2005: 339-340).

[1] For overviews of these attempts, see James (2011; 178-186), Kitcher (2006), and Wielenberg (2016).
[2] Philippa Foot (1967) was first to introduce this thought experiment.

Let us then consider the footbridge version of the previous case. Again, a runaway trolley is hurtling down a track and about to hit five people, who are tide-up and lying on the track. In this case, however, there is no lever that would enable us to divert the trolley. Instead, we are standing on a footbridge above the track, and we also aware that a stranger, who is very large in size, is standing on the same bridge. Let's also assume that the only way for us to stop the trolley is to push this stranger off the bridge. If we push him under the bridge, he will be killed by the trolley, but five lives will be saved at the same time. We can then ask: Are you permitted to push the stranger under the bridge? Interestingly, most people give a different response to this question compared to the previous case. Here, most people think that we are not permitted to push the stranger down from the footbridge in order to save the five (Greene 2016: 175; Singer 2005: 340).

Furthermore, most people do not merely give the previous intuitive responses to the previous cases, but their intuitive responses are also very strong and immediate (Greene 2002: 164). The previous kind of cases has led many philosophers to investigate whether they could find unified moral principles that would be able to not only explain but also justify our intuitive responses in those cases.[3] Yet, according to Peter Singer (2005), this is a wrong way to proceed – it is not the way to discover why we have the intuitions in those cases. Instead, he suggests that we have different intuitive responses to different cases because of the evolutionary origin of our moral intuitions (Singer 2005: 347-348).

[3] For a representative example, see Thomson (1976; 1985).

On this view, our moral intuitions are a product of natural selection.[4] In the distinct past, our ancestors presumably lived in hunter-gathering communities that consisted of very small social groups (Birdsell 1968: 232-233; Lieberman 2008: 168). In these circumstances, our ancestors needed to avoid violent personal acts within the community, including the acts of punching, pushing, kicking, hitting with sticks, and so on (Singer 2005: 347-348). If they failed to avoid these acts, this would have quickly led to a vicious cycle of revenge and escalation of violence within the small-sized community. As a consequence of this, the whole community could collapse due to social unrest. Hence, human beings evolved to have strong, immediate and emotion-based intuitive responses against any personal acts of violence, and so the act of pushing the stranger in the footbridge case triggers this sort of an intuitive response in us (Singer 2005: 347-348). In contrast, the act of pulling the lever in the original trolley case is not a personal violent act and therefore does not trigger a similar sort of a negative intuitive response in us, or so it has been argued.

This is why the previous two cases illustrate how we can often explain our moral intuitions by relying on their evolutionary origin. Furthermore, it seems that we can also explain the evolutionary origins of our moral beliefs more generally, or so I will suggest in this thesis. The objective of my thesis is then to address the question of whether the epistemic status of our moral beliefs is undermined by their evolutionary origin. In other words, my thesis will focus on the *epistemological evolutionary*

[4] I will further explain why this is the case in Chapters 2 and 4.

debunking arguments in meta-ethics (hereafter the EDA).[5]

Before I can explain what the EDA is, let me first outline the basic idea of the debunking arguments in general. Debunking arguments are genealogical arguments, which show that the epistemic status of our beliefs is undermined by their causal origin (Kahane 2011: 105-106). We can consider the following example of a debunking argument concerning non-moral beliefs and their causal origins. I have the belief that I went jogging this morning because I remember that I was jogging around the park earlier today. However, let us imagine that there is a pill that can make me believe that I went jogging and also that I suddenly discover that I took this belief pill just after I woke up. Although my belief can still be true (i.e., I really went jogging this morning), it can be argued that my discovery that I had the pill should undermine my confidence and faith in that belief.

Let us then return to the EDA. As a debunking argument, the EDA is basically a genealogical argument too. It too starts from an empirical claim according to which the evolutionary forces have had a significant influence on the moral beliefs and moral intuitions that we have. The argument then claims that, because of this evolutionary influence, our moral beliefs are not tracking the moral truth. The argument finally concludes that this fact undermines the epistemic status of our moral beliefs. The main purpose of this thesis is to develop this argument further and to defend it against several powerful objections that have been made to it recently.

[5] For overviews of different ways to understand the evolutionary debunking arguments, see Joyce (2016a: 142-152), Kahane (2011: 110-114), Leibowitz and Sinclair (2017: 210-211), Sinclair (2018: 98-99), and McPherson (2020: 30).

In the rest of this introductory section, I will outline the structure of my thesis. My thesis has two parts. **Part I** will examine how the EDA can be best formulated. In recent times, many philosophers have frequently discussed the EDA.[6] Nevertheless, there are three essential questions, which the evolutionary debunkers will need to be able to address, but currently there is a lack of unified and uncontroversial answers to all of them. Hence, the first part of my thesis on the EDA will try to answer the following three essential questions properly:

1st Question: How has evolution influenced and shaped our moral beliefs?

2nd Question: If evolution has shaped and influenced our moral beliefs in the way in which the true answer to 1st Question suggests, does this also mean that our moral beliefs are not tracking the moral truth reliably?

3rd Question: If our moral beliefs are not tracking the moral truth reliably, does this also mean that the epistemic status of our moral beliefs is undermined?

By answering to the three previous questions, my research will offer a concrete and precise way of understanding the EDA. Hence, **Part I** consists of three chapters and will address these three questions one by one.

[6] For overviews of epistemological evolutionary debunking arguments, see Nichols (2014) and Vavova (2015). Representative defenders of the epistemological evolutionary debunking arguments include Fraser (2014), Joyce (2006: 179-220; 2016b), and Street (2006; 2008).

Chapter 2 of my thesis will address the first question above. The evolutionary theory is widely accepted today. However, neither the proponents nor opponents of the EDA have so far provided an explicit and detailed explanation of the relationship between our moral beliefs and evolution. Hence, **Chapter 2** will describe how evolution has significantly influenced and shaped our moral beliefs. As a result, I argue that the origin of our moral beliefs can be wholly explained on the basis of evolution (or more precisely, on the basis of adaptation and exaptation).

Chapter 3 will then address the second question above. According to the EDA, if evolution has shaped and influenced our moral beliefs, our moral beliefs are not tracking the moral truth reliably. However, what does 'truth-tracking' actually mean? The point of **Chapter 3** is to offer two different theories of what kind of alignment between moral beliefs and moral facts truth-trackingness could consists. In **Chapter 3**, I will also argue that most, if not all, of our moral beliefs cannot be tracking the moral truth because their origin can be wholly explained on the basis of evolution.

Chapter 4 will address the third question. In this chapter, I will explain why the epistemic status of our moral beliefs is undermined as a result of the fact that these beliefs are not tracking the moral truth. I will also argue that, when we are aware of the evolutionary origin of our moral beliefs and also of the fact that our moral beliefs are not tracking the moral truth, the epistemic status of our moral beliefs is undermined due to both undercutting defeaters and higher-order defeaters. In addition, at the end of **Chapter 4**, I will also argue that, if externalism about justification were true, then our moral beliefs would have never been justified in the first place, given their unreliable

evolutionary origin.

After the discussion of how the EDA can be best formulated, **Part II** of this thesis will focus on two recent objections to the EDA and also on how the evolutionary debunkers can respond to both of these objections. There are, of course, many different objections to the EDA.[7] However, it seems to me that the two objections that I will address in **Chapter 5** and **Chapter 6** are the strongest ones. In other words, I believe that evolutionary debunkers will need to be able to deal with these two objections if they are to have any hope of making their argument widely accepted. In **Part II** of my thesis, I will also provide compelling responses to both of these objections in order to defend the EDA.

Firstly, **Chapter 5** will address the so-called 'conceptual truth objection'. The proponents of this objection argue that our moral beliefs can be justified on the basis of conceptual analyses on normative terms and also on the basis of moral propositions that are conceptual moral truths. They then argue that, as a consequence, the evolutionary forces are unlikely to have any debunking influence on the way in which our moral beliefs are justified. Nevertheless, in **Chapter 5**, I will outline two plausible ways in which the evolutionary debunkers can respond to the previous objection.

Finally, in **Chapter 6**, I will explore the second widely discussed objection to the EDA, that is, the so-called 'third factor objection'. The defenders of this objection concede

[7] For representative objections to the evolutionary debunking arguments, see Bogardus (2016), Copp (2008), Das (2016), Enoch (2010; 2011: 151-184); FitzPatrick (2014; 2015), Mogensen (2015; 2016), Shafer-Landau (2012), Vavova (2014), and Wielenberg (2010; 2014: 134-177).

that evolutionary forces have significantly influenced and shaped our moral beliefs. Nevertheless, they argue that certain third factors are correlated with both the moral truth and our evolutionarily shaped moral beliefs. Hence, according to the objection, the fact that there is an indirect correlation between the moral truth and our moral beliefs via the third factor is sufficient to ensure that our moral beliefs have been attained in a reliable way. The proponents of this objection argue the epistemic status of our moral beliefs cannot be therefore undermined by the relevant evolutionary considerations.

In **Chapter 6**, I will introduce three well-known versions of the third factor objection suggested by Copp (2008), Enoch (2010; 2011: 151-184) and Wielenberg (2010; 2014: 134-177; 2016) respectively. Furthermore, I will also argue that all these versions of the third factor response are too problematic for different reasons. As a result, I will conclude that there does not seem to exist plausible versions of the third factor objection, and so the EDA can be defended against this objection too.

To sum up, in this thesis, I will conclude that the EDA is a sound argument, and as a result, the epistemic status of our moral beliefs is undermined accordingly. Moral realists, particularly non-naturalist realists, thus face a serious epistemological challenge, and it is unclear whether they are able to deal with it given what I will argue in this thesis.

Part 1

The Evolutionary Debunking Argument

Chapter 2

The Evolutionary Origin of Our Moral Beliefs

2.1 Introduction

When it comes to how the EDA should be formulated, there really is no dispute over its starting point. All versions of the EDA begin from the idea that evolution has had a certain influence on the moral beliefs that we have, given that the evolutionary theory is, generally speaking, widely accepted to be true. Nevertheless, different philosophers who investigate the EDA have different views of what exactly this evolutionary influence on our moral beliefs has been like. The aim of this Chapter 2 is to explain how exactly evolution has influenced and shaped our moral beliefs.

Different views of how evolution shaped our moral beliefs can be classified under three very general categories. Firstly, most proponents of the EDA, if not all, argue that the evolutionary influence on our moral beliefs has been a "distorting" or an "illegitimate" one (Kahane 2011: 115; Street 2006: 109; Vavova 2015: 107). Secondly, many opponents of the EDA have also conceded that that evolution has influenced our moral beliefs at least to a certain extent. Nevertheless, they attempt to argue that the evolutionary influence on our moral beliefs has not been distorting or illegitimate (Enoch 2011: 168-174; Wielenberg 2014: 134-176). Finally, some philosophers have also attempted to reject the claim that our moral beliefs have been significantly influenced by the evolutionary forces (Huemer 2016; Parfit 2011: 534-542).

However, there is an important question that needs to be answered first before we can even begin to consider whether the evolutionary influence on our moral beliefs has been a distorting or an illegitimate one. That is, we first need to know *how* evolution, *as a matter of fact*, has influenced and shaped our moral beliefs before we can speculate about whether we ended up having distorted moral beliefs through the evolutionary process. Therefore, in the rest of this Chapter 2, I am going to introduce and evaluate three different theories of how evolution could have influenced and shaped our moral beliefs. I call these three views (i) the adaptation account, (ii) the exaptation account, and the (iii) cultural evolution account.

This chapter consists of four parts. Firstly, in §2.2, I will introduce three essential features of moral beliefs: (i) Moral beliefs have certain core contents that are commonly shared across different communities; (ii) moral beliefs are reliably connected to motivation; and (iii) moral beliefs are connected to our practices of praise and blame. In that sub-section, I will argue that any plausible view of the evolutionary origins of our moral beliefs must be able to explain how our moral beliefs came to have those three essential features.

Then in §2.3 and §2.4, I will introduce two views, the Adaptation Account and the Exaptation Account, of how evolution could have given our moral beliefs those three essential features. These two views cannot be both true because our moral beliefs cannot be both adaptations and exaptations at the same time. Nevertheless, in this thesis, I will remain neutral between whether the Adaptation Account or the Exaptation Account is true. Instead, as I will argue in the next Chapter 3, evolutionary debunkers can commit

to either one of these two views of how evolution influenced our moral beliefs. Hence, in the §2.3 and §2.4, I will focus on how these two accounts could be formulated in the best possible ways and I will also motivate these two accounts by explaining how they both have their own independent plausibility.

Finally, in §2.5, I will introduce the Cultural Evolution Account, which is the view that our moral beliefs are a result of cultural evolution rather than biological evolution. In that sub-section, I will point out that our moral beliefs can be explained at two different levels – the levels of ultimate causes and proximate causes. Moreover, I will argue that, in all the discussions of the EDA, what we really need to consider is the ultimate explanations of our moral beliefs instead of their proximate explanations. I will then argue that the proponents of the Cultural Evolution Account will have to rely on exactly the same ultimate explanations of our moral beliefs as the Adaptation Account and the Exaptation Account (even if they will defend a different kind of proximate mechanisms). As a result, in §2.5, I will conclude that evolutionary debunkers can simply put aside the Cultural Evolution Account.

2.2 Three Essential Features of Moral Beliefs

How exactly did evolution shape and influence our moral beliefs? Before I can provide an answer to this question, let us first consider the nature of those beliefs. Just like the case with most of the philosophically interesting notions, it is very difficult to define clearly what moral beliefs really are (McKay & Dennett 2009: 493). Nevertheless, I will next attempt to draw a very rough outline of the nature of moral beliefs. I suggest that moral beliefs have the following three key characteristics:

(1) Moral beliefs have certain core contents that are common across different communities.

(2) Moral beliefs are reliably connected to motivation.

(3) Moral beliefs are also connected to our practices of praise and blame.

In the rest of this sub-section 2.2, I will briefly explain the previous three features of moral beliefs. I will then investigate in sub-sections 2.3 and 2.4 whether the Adaption Account and the Exaptation Account could explain how our moral beliefs came to have those three features.

2.2.1 Feature 1: The Core Contents of Moral Beliefs

The first feature is that moral beliefs across different communities share certain core contents. For instance, most of us believe that we should help and co-operate with others, that we should keep our promises and not tell lies, that it is wrong to harm and kill other people, and so on. These very conventional moral beliefs are commonly shared by the members of most communities and societies.[8]

Moreover, the previous core contents of moral beliefs are also manifested in the conventional moral practices of most communities and societies.[9] For example, most

[8] This feature of moral beliefs is similar to the idea of "moral fixed points" suggested by Cuneo and Shafer-Landau (2014). I will further discuss the moral fixed points theory in Chapter 6.

[9] For further discussions of this view, see Huemer (2016: 1988-1994) and Mesoudi and Danielson (2008:

of us think that it is wrong to punch your friends in the face and so we won't easily do so. Likewise, we think that it is good to take care of our family members, which is why we take our parents to the hospital when they are sick. These ordinary moral practices and habits are commonly shared by most, if not all, communities and societies, and the core contents of our moral beliefs are manifested in our ordinary moral practices and habits.

However, it is worthwhile to remember that the fact that our moral beliefs have these core contents does not mean that they are thus justified or true. For example, the moral belief that slavery is morally permissible was part of the conventional moral practices of most communities and societies in the past. However, even if this moral belief was widespread in many societies, this does not imply that it was thus justified or true.[10]

2.2.2 Feature 2: Moral Beliefs are Reliably Connected to Motivation

Another essential feature of moral beliefs is their practicality (Smith 1994: 6-7). It is widely accepted that there is a close connection between our moral beliefs and what we are motivated to do (Gibbard 1990: 110; Street 2006: 159; Vahid 2009: 1). In other words, if we hold a given moral belief, we are in most cases motivated to act in accordance with that belief. This suggests that there is a reliable connection between moral judgments and being motivated to act accordingly.

Let us consider the following example inspired by Michael Smith (1994: 6-7). Assume

230).

[10] For a further explanation of why the fact that some moral beliefs are widespread in many societies does not imply that they are thus justified or true, see Copp (2001: 9-36).

that Marina does not initially have the moral belief that it is good to help other people in need. After you discuss the issue with her, she claims that she has now changed her mind and thus she now comes to acquire the previous moral belief. In this case, we expect that Marina would be motivated to act according to her new moral belief. Imagine that, after Marina's claim about her change of mind, you and she meet a volunteer collecting money for the local charity for homeless people. You will expect Marina to donate some money to the charity if she can easily afford to do so.

However, if she claims that she has no desire to donate any money at all, you will begin to doubt whether she really changed her mind (Smith 1994: 6-7). This is because, if she really acquires the previous moral belief, we will expect her to be motivated to act according to that moral belief. In this example, our expectations of Marina being motivated to act illustrates nicely the second key feature of moral beliefs: Moral beliefs are assumed to be reliably connected to motivation. Having moral beliefs seems to imply that we have corresponding motivation to act accordingly.[11]

2.2.3 Feature 3: Moral Beliefs are Connected to Our Practices of Praise and Blame

The third essential feature of moral beliefs is that they are connected to our practices of blaming and praising others. Most moral beliefs are related to whether a given act or a

[11] I aware that there is a debate between moral judgments internalism and externalism. Both internalists and externalists accept that there is a reliable connection between moral judgments and motivation. However, they hold different views of explaining this connection. It is beyond the scope of this thesis to discuss further the question of which view is true, but see Brink (1989: 37-80), Smith (1994: 60-91), and Svavarsdottir (1999). In this thesis, I will just defend the view that moral beliefs are assumed to be reliably connected to motivation, and I believe that neither internalists nor externalists would disagree with this idea.

certain behaviour is good, bad, right or wrong. Moreover, some philosophers suggest that, whenever there is a morally wrong act X, any person who performs this act X also deserves blame. For example, in his book *Utilitarianism*, J.S. Mill (1861/2003: 222) explicitly suggested that:

> We do not call anything wrong, unless we mean to imply that a person ought to be punished in some way or other for doing it; if not by law, by the opinion of his fellow creatures; if not by opinion, by the reproaches of his own conscience.

Thus, according to Mill, whenever we think that a certain act is wrong, we also think that a person who acts in that way deserves to be punished, at least in some way. This punishment can mean different things in different cases. In some serious cases, the person ought to be punished by law, whereas perhaps in some less serious cases she ought to be only blamed by her conscience or the opinions from others.

After Mill, many philosophers have held similar but weaker views of the relationship between moral wrongness and blameworthiness. These philosophers firstly tend to distinguish blameworthiness from any legal punishments. They thus set aside the question of whether there should be a reliable relationship between moral wrongness and legal punishments (Gibbard 1990: 41).[12] Secondly, and more importantly, they argue that an agent can perform a morally wrong act even if she is not blameworthy for

[12] Allan Gibbard (1990) introduces a very good example, namely parking, to explain why legal punishment does not necessarily imply moral wrongness (41). We all think that a fee should be charged for failing to pay a sufficient fee for parking, but we do not think that doing so is morally wrong. Also, we normally would not feel guilty if we did so ourselves.

acting in that way. For example, Brad Hooker (2002) suggests that an agent can act in a certain way because she has "false beliefs about the circumstances" (73). Consider an example: Diego was approaching Jolene, and Jolene then attacked Diego because she thought he was a violent hooligan. Unbeknownst to Jolene, Diego is actually an innocent passer-by who just happened to look like someone else. In such a case, Jolene's action was based on a false belief. Thus, in this case, if Jolene is not responsible for her false belief, she can have a good excuse for what she did, and thus we should not blame her for acting in that way. Nevertheless, what she did was still morally wrong, as it is morally wrong to attack an innocent person.

In this thesis, I will accept the weaker view of the relationship between moral wrongness and blameworthiness. According to this view, there is a reliable relationship between moral wrongness and blameworthiness: When an agent performs an act that is morally wrong, *by default and other things being equal* she ought to be blamed for acting in that way.[13] But, in some cases, we should not blame an agent for acting in a morally wrong way if she can be excused for acting in that way, just like the previous example illustrated (Hooker 2002: 73).[14]

[13] It is difficult to explain this relationship between moral wrongness and blameworthiness explicitly. This is because there are at least three different proposals of what this relationship could exactly be, suggested by Copp (2001: 25-26), Gibbard (1990: 41-47) and Hooker (2002: 73-75) respectively. Moreover, it seems that the defenders of different ethical theories would defend different proposals of the relationship between moral wrongness and blameworthiness. For example, Brad Hooker (2002) defends rule-consequentialism and he also suggests that a plausible account of moral blameworthiness should be somehow related to the consequences of actions (72-75). In contrast, Allan Gibbard (1990) seems to rely on certain assumptions and theories of the relationship between rationality and morality to explain the relationship between moral wrongness and blameworthiness (41-42). Therefore, it is beyond the scope of this thesis to evaluate all three proposals. What I just want to emphasize is that all three proposals share the same core idea: There must be a reliable relationship between moral wrongness and blameworthiness.

[14] For a further discussion of blameless wrongdoing, see Parfit (1984: 30-34).

Following the same line of this reasoning, I will also accept a similar view of the relationship between moral rightness and praiseworthy: When an agent performs an act that is morally right, *by default and other things being equal* she ought to be praised for acting in that way. But in some cases, we should not praise an agent for acting in the morally right way. For example, the agent ought not to be praised if she accidentally performs such a morally right action.

2.3 1st Theory of Explanation: The Adaptation Account

In this §2.3, I will introduce a plausible view of how evolution could have influenced our moral beliefs to have the previous three essential features, that is, the Adaptation Account. According to this account, having moral beliefs with the three essential features could be thought of as an adaptation. Thus, this §2.3 consists of two parts. Firstly, in 2.3.1, I will explain the role that adaptation is generally thought to play in evolution. Secondly, in 2.3.2, I will introduce why having moral beliefs with the three essential features described in the previous sub-section 2.2.2 could be considered to be an adaptation.

2.3.1 Adaption in Evolution

In this 2.3.1, I will introduce the basic crux of what role adaptation plays in natural selection according to the evolutionary theory.[15] We can begin by asking the basic question: What is an adaptation? Firstly, it must be emphasized that "the status of

[15] There are many interesting questions related to adaptation. For instance, what is the role that genes play in adaptation and natural selection (Orr 2005)? What is the relationship between long-term and short-term measures of adaptedness (Brandon 1990: 24-27)? Nonetheless, I will put aside these questions as it is beyond the scope of this thesis to discuss them in detail.

adaptation always belongs to traits" (Clarke 2018: 35). Very roughly, an organism's traits include its physical features, behavioural tendencies and other phenotypic characteristics (i.e., the characteristics that distinguish the organism from other species in recognizable ways).

Let us consider salmons as an example. Firstly, salmons have certain physical features. They have gills, scales and flesh that is orange to red in colour. Secondly, salmons also have certain behavioural tendencies. For instance, they are anadromous: They are born in fresh water (usually rivers) and then migrate to sea. They also return to fresh water to spawn and to reproduce. Finally, some of these physical features and behavioural tendencies of salmons can be considered to be phenotypic characteristics (or simply *phenotype*s). *Phenotypes* are the traits that make salmons distinguishable from other fishes and organisms in a recognizable manner (Brandon 1978: 198). We can distinguish these *phenotypes* by direct inspection, measurement or description (Johannsen 1911: 134). For example, we can observe that salmons have flesh of a certain colour (usually orange), and we can also describe the salmons' anadromous behavioural tendencies. The colour of flesh and their anadromous behaviour are examples of the *phenotype*s that make salmons distinguishable from other species of fish.

Although adaptations are all traits, not all traits are adaptations. This is because adaptations must have two essential features. Firstly, an adaption is a trait that can serve a certain purpose or a function in a teleological sense (Clarke 2018: 38; Killin 2018: 378). Let us take polar bears and their heavy fur as an example. Polar bears' heavy fur

can serve the function of keeping their body warm (Clarke 2018: 38). Moreover, their fur has this function because polar bears and their ancestors have been living in cold climates. Therefore, a polar bear's fur is an adaption, which serves a certain function and a purpose in the natural environment in which the animal lives.

Let us then consider the second key feature of adaptations. It is thought that being an adaptation has a close relationship with natural selection. Let me first briefly explain how natural selection works. According to the standard interpretations of evolutionary theory, natural selection is not a single step process, but rather there are (at least) three steps in any process of natural selection. The first step is *variation* (Brandon 1990: 7). We can consider the example in which organisms within a particular reproducing population initially had almost exactly the same traits. Some of these organisms then acquire new traits as a result of a random change in their genes, or in other words, genetic mutation.[16] *Variation* thus consists of the divergence among organisms within the reproducing population due to the differences in their traits and genes.

The second step is *heritability*. As I have just mentioned, variants within the reproducing population differ in the traits that they possess as a consequence of genetic mutation. In cases of natural selection, we only consider the new traits that are *heritable*. This is because natural selection is about "evolutionary" and "cross-generational" change (Brandon 1990: 6). If a given trait of the variants is *heritable*, then the offspring of these variants (in succeeding generations) will also have the same trait due to the

[16] For example, different kinds of snakes came to have different kinds of venoms as a consequence of genetic mutation (Carlin 2011).

genes that they have inherited from their parents.

Finally, the third step of the process is *selection*. The definition of this step can be roughly put as follows: Natural selection is about differential reproduction due to differential fitness within a common environment (Brandon 2005: 160). That is to say that different variants with different traits will "leave different numbers of offspring in succeeding generations" within a common environment (Brandon 1990: 7). These three steps together provide us with the basic idea of how the natural selection works.

We can then turn to the relationship between adaption and natural selection. As Ellen Clarke (2018) suggests, "[a]daptation entails that there has been [natural] selection" (37). Some variants leave more offspring than other variants. According to many evolutionary theorists, this is because they are *better adapted* than other variants within the common environment (Brandon 1978: 182; Darwin 1876/2009: 62-68). The organisms that are better adapted have certain traits with certain functions, which effectively enable these organisms to leave more offspring in the circumstances in which they live.

Let us consider polar bears and their thick fur as an example again. The ancestors of polar bears presumably lived with many other closely similar organisms within a reproducing population. Moreover, there was variation among the ancestors of polar bears and their relatives, which had to do with the trait of having thick fur. This feature had the function of keeping the animals warm. The ancestor of the polar bears also lived around the North Pole in a cold climate. In this situation, the ancestors who had thick

fur were better adapted to the environment than the other variants that did not have equally thick fur. This is because the thick fur could keep these animals warm and thus increased their chance of survival. Moreover, having this feature increased the chance of these animals to grow until they could reproduce as mature animals. As previously mentioned, an adaptation is thought to be a trait with a certain function that makes organisms leave more offspring in the circumstances that they are living in. This is why the thick fur of polar bears count as an adaptation.

2.3.2 Moral Beliefs as an Adaptation

In this sub-section, I will explain why having moral beliefs with the three essential features (as outlined in 2.2) could thought be an adaptation. Let us recall the three core features that our moral beliefs have:

i. Certain core contents that are shared across different communities
ii. A connection to motivation
iii. A connection to the practice of blame and praise

Why would having moral beliefs with these three features be an adaptation? As just explained, adaptations are traits with certain functions that help the organisms to leave more offspring in the succeeding generations. Therefore, for having moral beliefs with the previous three features to be an adaptation, having moral beliefs of that kind should have helped our ancestors to leave more offspring in the succeeding generations.

Thus, in the rest of this 2.3.2, I will explain why having moral beliefs with the previous

three essential features should have helped our ancestors to leave more offspring in the succeeding generations. Let us then start from the first feature: Moral beliefs have certain core contents that are commonly shared across different communities. Some examples of the core contents are that we tend to believe that helping others and co-operating with them is the right things to do (James 2011: 58-60; Joyce 2000: 714; Ruse 1995: 230-231; Sinclair 2012: 649-652). Now let us assume that our ancestors had moral beliefs with the previous core contents. Consider then the circumstances in which our distant ancestors lived: They lived in small societies that had very limited resources to defend themselves against any external threats. Unlike today, the networks of relationships of our distant ancestors were very small and close.

In these circumstances, as Scott James (2011) suggests, "an individual who could routinely count on … non-relatives for assistance – in return for giving assistance – would have possessed a pronounced advantage over an individual unable or unwilling to forge such relationships" (58-59). If our ancestors then came to have the moral beliefs that helping others and co-operation are the right things to do, it would have been much easier for them to live with others, for example, in the circumstances in which they faced limited resources and different natural disasters. Given that they really acted in the ways that corresponded to their moral beliefs that helping others and co-operation are the right things to do in these circumstances, they could avoid violent personal acts – including the acts of hitting and punching others – within the small-sized community (Singer 2005: 347-348). As a consequence of this, they could prevent a vicious cycle of revenge and escalation of violence within the community, and they could also avoid

the collapse of the whole community due to social unrest.[17] Hence, it is reasonable to think that the actions of helping others and co-operation would have helped our ancestors to leave more offspring.[18]

We can then focus on the second key feature of our moral beliefs. It would not have been sufficient if our ancestors merely came to have the previous moral beliefs with the relevant core contents. In addition, they had to be able to act in ways corresponding to those moral beliefs. If the moral beliefs of our ancestors were not reliably connected to what they were motivated to do, then those beliefs would have been unrelated to whatever these ancestors really did and thus unable to affect the reproductive rate of our ancestors. Hence, the relevant practical feature of moral beliefs – that they are reliably connected to motivation – would have helped to motivate our ancestors to act in the ways that were reproductively beneficial (i.e., to act in accordance with the moral beliefs with core contents).[19]

Finally, we can turn to the third key feature of moral beliefs. That is, moral beliefs are generally thought to be connected to our practices of blame and praise. To explain why moral beliefs with this feature are also an adaptation, let us consider two cases. Firstly, let us imagine that some of our ancestors had moral beliefs that only had the abovementioned two features. Secondly, let us also imagine that another group of our ancestors had moral beliefs that had all three features of moral beliefs, including being

[17] I also discussed why this is the case in Chapter 1.
[18] This claim has been widely discussed in evolutionary biology. See Hamilton (1963), Nowak (2006), and Trivers (1971).
[19] For further discussions of how our moral judgments can be thought to be connected to what we are motivated to do, see Gibbard (2003: 152-158) and Prinz (2015).

connected to the practices of blame and praise. The latter group of our ancestors would have left more offspring in the succeeding generations, and thus having moral beliefs that have the third feature is an adaptation, or so I will argue next.

The difference between these two groups of ancestors, I argue, is that the members of the latter group would have had an additional incentive to act in the evolutionarily advantageous ways due to the third, additional feature of their moral beliefs (For the sake of simplicity, in this example, I will merely consider moral beliefs with the feature of being connected to the practices of blame). When someone is being blamed for doing an act, this can produce in her an uncomfortable attitude that we usually call guilt or remorse (Brandt 1979: 166; Gibbard 1990: 126). It can be then argued that the prospects of guilt and remorse then provide an extra motive for agents not to perform certain acts that would make them have these negative self-directed emotions (Brandt 1979: 166-167).[20]

We can now return to the previous example of the two groups. Imagine that there are two selfish persons who suddenly join the previous two groups. Even if the members of the first group are motivated to co-operate with others, they are not particularly motivated to blame or punish the selfish newcomer when she does not help other people in the community. Moreover, they will not exclude this newcomer from cooperating with them either. As a result, we can imagine that this selfish person would be very

[20] Similarly, when a person is praised for doing a certain act, this could give her a positive and comfortable feeling that we usually call pride (Bennett 2016: 52-53). For instance, we usually get praised if we donate money to charity. Also, we often feel proud as a result of making charitable donations, and for this reason too we are more likely to perform similar actions in the future (Brandt 1979: 168). This is because the positive emotions provide an extra motive for us to perform the actions that we are likely to be praised for (Brandt 1979: 168).

successful in this society: She gets helped by others, but she does not need to help anyone else. Consequently, she would leave many offspring who presumably would also be selfish in nature. This society would then gradually become a more selfish society and thus a less successful society.[21]

In comparison, the selfish person who joined the second group would have a different fate because she would be blamed and sanctioned by others (Gaus 2011:105). As a result, she might also feel guilty, which would make her less likely to be selfish in the future. Furthermore, she would also get a bad reputation after which others would not be willing to co-operate with her. Scott James (2011) suggests that one's reputation must have been very important for our distant ancestors in their small communities (60). When our ancestors had the reputation of not helping others, their neighbours would have distanced themselves from them and, as a result, it would have been extremely difficult for those selfish ancestors to survive. For their own interests, our ancestors would have tried to avoid being blamed by others because they would not want to be excluded from co-operation in the community.

Hence, if our ancestors had the moral beliefs with the feature of being connected to the practices of blame, they could prevent their society from becoming a selfish society. Moreover, the members of a generally unselfish society would tend to help and co-operate with others. As I just mentioned above, it is reasonable to think that the actions of helping others and co-operation would have helped our ancestors to leave more

[21] For further explanations of why a selfish society would not be very successful, see Bowles and Gintis (2011: 19-78), James (2011: 58-61), and Richerson (2011). For a further clarification on the term 'selfish' in psychology and evolutionary theory, see Wilson (1992).

offspring. As a result, having moral beliefs that are connected to the practice of blame (and praise) is also an adaptation.

In this §2.3, I introduced the Adaptation Account of how evolution could have influenced our moral beliefs. I also explained why, according to the Adaptation Account, having moral beliefs with the three essential features is an adaptation. There is, however, another competing explanation of how evolution might have influenced our moral beliefs. In the next sub-section, I will introduce this competing theory of explanation, that is, the Exaptation Account.

2.4 2nd Theory of Explanation: The Exaptation Account

Evolution is not just about adaptations but also about exaptations. The notion of exaptation was firstly introduced by evolutionary biologists Stephen Jay Gould and Elisabeth S. Vrba (1982) in an article entitled "Exaptation – A Missing Term in the Science of Form".[22] Roughly speaking, a currently adaptive trait is thought to be an exaptation if this trait was either selected for a different function in the natural selection or is a by-product of another trait. Furthermore, according to the Exaptation Account, having moral beliefs with the three essential features is not an adaptation but rather an exaptation.

[22] One of the reasons why biologists introduced the term 'exaptation' was to replace a similar but problematic term "preadaptation" (Dennett 1995: 280). The latter is a problematic term because it seems to imply that there is a certain "predestined" function that a given trait has such that the organism can rely on it in future (Gould 1991: 144n). For a further discussion of the distinction between exaptation and preadaptation, see Dennett (1995: 280-281). Another term that is similar to exaptation is 'secondary adaptation' (Gould & Lewontin 1979: 596). For a further discussion of the distinction between exaptation and secondary adaptation, see Fraser (2010: 226).

In my thesis, I will remain neutral between whether the Adaptation Account or the Exaptation Account is true. This is because I believe that evolutionary debunkers can accept either one of these two views of how evolution has influenced our moral beliefs. In the next Chapter 3, I will explain why debunkers can argue that our moral beliefs are not tracking the moral truth reliably due to their evolutionary origin in a way that does not depend on whether the Adaptation or the Exaptation Account is true. Therefore, in §2.4, I will merely attempt to introduce the Exaptation Account and explain why it is also a plausible view of the evolutionary origin of our moral beliefs. This sub-section thus consists of two parts. Firstly, 2.4.1 will introduce what role the evolutionary theory generally takes exaptation to play in natural selection. Secondly, 2.4.2 will explain why having our moral beliefs could also be thought to be an exaptation.

2.4.1 Exaptation in Evolution

In this 2.4.1, I will introduce what role the evolutionary theory generally takes exaptation to play in natural selection. It should be noted that exaptation is not an alternative concept that is used to replace adaptation. Instead, it is considered to be an additional element in the explanation of what happens in natural selection. This is because many natural phenomena cannot simply be explained as adaptations and so the evolutionary explanations of these phenomena must also rely on exaptations.

We can begin by returning to the idea of adaptation in natural selection. Whenever we suggest that a certain trait is an adaptation, we are describing why this trait has been selected for and thus why it is widespread within a particular reproducing population (Schulz 2013: 195). For example, polar bears' heavy fur has been adaptive for its

function of keeping their body warm and thus has been selected for this function. Nevertheless, it is possible that an adaptation can be no longer adaptive later on. For instance, human appendix is considered to be an adaptation because it served a certain function for our distant ancestors in the past, but it is no longer adaptive for the original function today (Sterelny & Griffiths 1999: 217-218).

In contrast, if a trait is thought to be an exaptation, it must be currently adaptive as a result of having a certain function (i.e., this function effectively enables the organisms which currently have the trait to leave more offspring). Yet, this trait did not serve its current function when the organisms came to have the trait. According to Armin Schulz (2013), the notion of exaptation can thus be defined as follows (196): A trait *T* is an exaptation if and only if (1) *T* is adaptive for a certain function *F*, and (2a) *T* was selected for a different function *F'* or (2b) *T* is a by-product of a different adaptation *T'*.[23]

Let us then consider exaptations understood in the way just outlined. For instance, the feathers of birds currently serve a function that is useful for flying, but this trait was originally evolved for "conserving temperature" (Ayala 2010: 9019). Nevertheless, this trait then later on gets co-opted to serve the additional function of flying (Ayala 2010: 9019; Brandon 1990: 172n14). This example shows how the new function of an exaptation can co-exist with the original function, as the feathers of birds currently

[23] The notion of exaptation actually entails a historical view of functions (Godfrey-Smith 1994; Sinclair 2012: 642). For an explanation of the historical approaches to functions, see Wright (1973; 1976). For an explanation of how this view of functions can be related to the notion of exaptation, see Godfrey-Smith (1994: 357-359). In contrast, for an overview of the ahistorical theories of functions, see Cummins (1975). For an objection to the ahistorical theories of functions, see Garson (2019).

serve both functions of conserving temperature and also flying.

Let us consider the lungs of mammals as another example of an exaptation that was originally selected for a different function. According to some biologists, lungs were originally evolved in the form of swim bladders to "aid predatory [fishes] in pursuing prey" (James 2011: 14; Darwin 1876/2009: 147-148). Some ancestors of these fishes then began to live on land. It turned out that those bladders were "well suited to respiration", and so they eventually became lungs (James 2011: 14). Even if at one stage the lungs were used by our ancestors to pursue prey in the water, lungs currently serve a different (adaptive) purpose for mammals. This example thus illustrates how the new function of an exaptation can also sometimes replace the older function that it was originally selected for.

Finally, if an exaptation is a by-product or a "spandrel" of another adaptation, then it does not even need to have any original function at the time when it evolved (Joyce 2017: 102). Consider hearts as an example. It is unlikely that the beating sound of human hearts evolved to serve any purpose (indeed, it could have been even better if humans could have circulated blood silently) (Brandon 1990: 42). This trait is therefore considered to be a by-product of human blood circulation. Nevertheless, it is very common today for doctors to diagnose patients by listening to the beating sound of their hearts. As a result, the beating sounds of human hearts is an exaptation because it now serves a certain purpose that is beneficial to us (Brandon 1990: 42). But this exaptation did not have any original function when it evolved as it is just a by-product of another adaptation.

2.4.2 Moral Beliefs as an Exaptation

The Exaptation Account argues that our moral beliefs are an exaptation instead of an adaptation. According to the previous definition of exaptation of an exaptation, A trait *T* is an exaptation if and only if (1) *T* is adaptive for a certain function *F*, and (2a) *T* was selected for a different function *F'* or (2b) *T* is a by-product of a different adaptation *T'*. In the previous 2.3, I already explained why having moral beliefs with the three essential features could be thought to be adaptive. In order to argue that our moral beliefs are an exaptation instead of an adaptation, the Exaptation Account thus needs to argue that these beliefs were either selected for a different function or are a by-product of another adaptation.

Below, I will suggest that if our moral beliefs were an exaptation, they would be an exaptation because they would be a by-product of another adaption. I will suggest that our moral beliefs could be thought to be a by-product of our emotional systems. Hence, 2.4.2 consists of two parts. Firstly, I will explain why our emotional system is not merely required for us to be able to have moral thoughts, but rather it also imposes certain constraints on which moral beliefs we can have. Secondly, I will also explain how moral beliefs with the three essential features could also be argued to be a by-product of our emotional systems and thus how these beliefs could also be understood as an exaptation.

(i) Emotional system and ability to think morally

Many philosophers and biologists have suggested that morality in general is a by-

product of other non-moral faculties and abilities that we have evolved to have (Ayala 2010). But what exactly are these faculties and abilities that we first evolved to have? I will briefly introduce two proposals next.

The first proposal is from Jesse Prinz (2006; 2007). Prinz (2007) suggests that ethical behaviour and the capacity to moralize are by-products of our emotional systems (such as anger, disgust, shame, and so on) (263). Another potential proposal was already suggested by Charles Darwin (1871/2009), the founder of the evolutionary theory (71-72). According to Darwin (1871/2009), moral sense and conscience are a by-product of two kinds of non-moral capacities (71-72). Firstly, we need to have certain advanced intellectual abilities, such as our ability to use language (Darwin 1871/2009: 71-72). Secondly, we also need to have certain social instincts, including our feelings of sympathy for others in our society (Darwin 1871/2009: 72).

I will put aside the debate about which non-moral capacities were responsible for giving us our moral capacities as a by-product.[24] In the rest of this sub-section, for the sake of simplicity, I will merely focus on how our capacity to think morally could be thought to be the by-product of our emotional capacities (I will also discuss how our moral capacities could be thought to be the by-product of our general intellectual capacities elsewhere in Chapter 5).

[24] Although these two proposals suggested by Prinz and Darwin respectively are very different, they are not incompatible with each other. In other words, it is possible that moral capacities are actually a joint by-product of the non-moral capacities suggested by both proposals. Nevertheless, I am more inclined to think that moral capacities are the by-product of our emotional systems. Unfortunately, it is beyond the scope of this thesis to further investigate this interesting question.

I will start from the idea that our emotional systems could be thought to be necessary for our moral thought. Let us consider the following example of an android robot (Tiberius 2015: 69-70; Prinz 2006: 29-32). Imagine that a scientist has invented an android robot who has self-consciousness and also other higher-level cognitive capacities. Moreover, this robot knows every moral theory in the world, including utilitarianism and deontology. Nevertheless, this robot lacks emotions and hence never feels sad, angry, happy, guilty and so on.

Let us imagine that we then ask the robot: What do you think about the case where an innocent person is tortured by a group of hooligans? The robot should be able to conclude that these hooligans are doing something wrong because it knows that torturing an innocent person largely diminishes utility and violates the Kantian categorical imperative. Nevertheless, the robot would not feel any disgust or anger when it thinks of the wrongful actions of the hooligans. In this case, it seems that the robot is actually unable to think morally. This is because the robot fails to be disposed to have any negative emotions towards morally wrong actions and the persons who perform them. In contrast, we can also consider a human being who knows nothing about any moral theory, but she has an extremely strong negative emotion towards torture and what the hooligans are doing. Intuitively, we think that this person, instead of the robot, is able to think morally (Prinz 2006: 32).

The previous example thus suggests that emotional capacities are required for us to be able to think morally. The fact that our emotional systems are necessary for us to think morally, however, does not imply that our moral beliefs are necessarily a by-product of

our emotional systems. Instead, if our moral beliefs with the three essential features are a by-product of our emotional systems, then there would have to be some constraints imposed by the emotional systems on what moral beliefs we can have.[25] The proponents of the Exaptation Account would thus need to identify these constraints in order to be able to argue that our trait of having moral beliefs with the three essential features is a by-product of our emotional systems.

To illustrate the way in which our moral beliefs could be thought to be constrained by our emotions, we can consider the following example of sibling incest famously suggested by Jonathan Haidt (2001: 814). In Haidt's example, a pair of siblings decided to have sex only once. They used many reliable forms of birth control when they had sex, and afterwards they decided to keep their sexual activity absolutely secret. In Haidt's experiment, most subjects intuitively and immediately thought that the siblings were doing something wrong. Nevertheless, when most subjects were further asked to explain why they thought this, they could only suggest the reasons that were irrelevant to the case as they were already ruled out by the vignette itself. For example, most subjects first argued that the siblings might have a disabled child as a consequence of inbreeding. But, they were then reminded that the siblings used many very reliable forms of birth control. When further asked why they thought that the siblings were

[25] Consider whales' tail-flukes as a biological example. Whales' tail-flukes are horizontal instead of vertical, and there is no significant advantage for whales to have horizontal tail-flukes. Andreas Mogensen (2016) suggests that whales have horizontal tail-flukes because they inherited this structure from their ancestors who had not been living in water (1813). Whales' tail-flukes are a kind of phylogenetic constraint imposed by their past adaptations that originally served other functions. Even though the whales currently use their horizontal tail-flukes to swim, "the [phylogenetic] constraint itself was not eliminated" (Eastwick 2009: 797). In contrast, fishes' caudal fins are vertical, and there are no similar phylogenetic constraints imposed on fishes' caudal fins by their past adaptations. As a result, fishes do not need to dodge any phylogenetic constraints to swim with their adaptive caudal fins.

wrong to have sex, most subjects then admitted that they did not conclude that the siblings did something wrong on the basis of moral reflection or deliberation (Haidt 2001: 814). Instead, the relevant research suggests that they just came to have this moral belief because they found the sexual relations between siblings disgusting, and they tried to find some reasons to support their belief afterwards (Prinz 2006: 30-31).

By using this example and others like it, Haidt (2011) argues that many of our moral beliefs are caused by our quick moral intuitions that consist of "affective valence (good-bad, like-dislike)" (817-818). Moreover, this affective valence has a clear emotional origin: Most people form the moral belief that the siblings were doing something wrong merely based on their negative emotion of disgust towards sexual relations between siblings. As a result, our moral beliefs in this case can be argued to be produced in a way that is constrained and influenced by our emotions.[26]

(ii) Moral beliefs as a by-product of emotional systems

By using Haidt's example of siblings' incest, I have explained why our moral beliefs can be argued to be produced in a way that is constrained and influenced by our emotions. In the rest of this sub-section, I will further explain the additional reasons in support of this conclusion.

[26] Wheatley and Haidt (2005) also conducted another experiment to investigate how our negative emotions can lead us to form certain moral beliefs. In the experiment, the hypnotized subjects were asked to feel disgust when they heard the word "often" (Prinz 2006: 31). They were then presented with vignettes that contained the word 'often'. Moreover, some of these vignettes described morally admirable characters. Nevertheless, many hypnotized subjects judged that the morally admirable characters described in those vignettes were morally bad. This is because those subjects had negative emotions when they heard the word 'often' in the vignettes. Consequently, this experiment shows that our moral beliefs could be influenced and constrained by the affective valence that has an emotional origin.

Consider the first feature: Our moral beliefs have certain core contents that are shared across different communities. The defenders of the Exaptation Account can argue that much of these core contents can be explained by the emotional systems that all human beings have evolved to have. In order to see this, let us return to the example of two versions of the trolley problem that I have also mentioned in the introductory chapter. There, I already explained that most people have different intuitive responses to the classic and the footbridge version of the trolley problem. According to some philosophers and psychologists, this is because the action of pushing someone to her death in the footbridge case is "more emotionally salient" than pulling a lever in the classic trolley case (Greene et al. 2001: 2106).

In the footbridge case, pushing someone down from the footbridge can be considered to be a personal violent act, just like the acts of punching, pushing, kicking, hitting with sticks, and so on. We typically have immediate and negative emotional responses towards such personal violent acts, and the act of pushing the stranger off the bridge will trigger this sort of our intuitive responses (Singer 2005: 347-348). In contrast, in the classic trolley problem, the action of pulling a lever does not resemble a personal violent act, and therefore it will not trigger our negative emotional responses. Such a difference in our emotional responses towards the actions of pushing someone off the bridge and pulling the lever has also been empirically proven in an MRI scan experiment (Greene et al. 2001: 2106-2107).

This example of the two versions of the trolley problem shows that our moral beliefs can be claimed to be caused by quick affective moral intuitions. Moreover, we also have

many negative attitudes, such as disgust and anger, towards people who perform violence personal acts and harm others. These negative attitudes could be argued to ground or give rise to our quick moral intuitions of the wrongness of these actions. Hence, the reason why we hold our moral beliefs with the relevant core contents (for example, wanton killing is wrong, that it is wrong to rape, and so on) could be argued to be that we are likely to be disposed to have negative emotions towards these actions involved in the core contents. And, these negative attitudes can be thought to ground and give rise to our quick moral intuitions of the wrongness of these actions of which the core contents of our moral beliefs are about. As a result, having moral beliefs with certain core contents seems to be a constraint imposed by our emotional systems.

We can then consider the second feature of our moral beliefs: Moral beliefs are often thought to be reliably connected to motivation. If our moral beliefs are produced in a way that is constrained by our emotions, then we can easily explain why there is a reliable connection between our moral beliefs and motivation. It is widely accepted that our emotions have strong motivational effects (Kauppinen 2015; Prinz 2015). For instance, if you see a pile of vomit on the street and you find it disgusting, you will use your hand to cover your nose and also walk away from the pile of vomit as soon as possible.

Let us then consider the example of moral beliefs. As previously mentioned, we have negative emotions towards torture. These negative attitudes do not merely give rise to our quick moral intuitions that cause us to believe that torturing is wrong, but rather we are also motivated not to torture others because of these negative emotions toward the

acts of torturing. As a result, if our moral beliefs were constrained by our emotional systems, this would also explain why these beliefs are reliably connected to the motivation.[27]

Finally, moral beliefs with the third feature could also be argued to be produced in a way that is constrained by our certain emotions. As we saw above, moral beliefs are often thought to be connected to our practices of blaming and praising (i.e., the third feature). Moreover, the practices of blaming and praising are closely related to certain reactive attitudes, which are the feelings and emotions that we naturally have when we are involved in inter-personal human relationships (Strawson 1962/2008: 9-10).

Let us consider an example where Carrie broke a promise. When she is being blamed for breaking the promise, this can normally produce in her an uncomfortable attitude that we call guilt and also in others an attitude that we usually call resentment towards Carrie (Gibbard 1985: 9-11; Strawson 1962/2008: 7). Likewise, when a person is being praised for acting in a certain way, this often gives her a positive reactive attitude that we call pride and produces in others attitudes that we call gratitude or admiration towards that person (Bennett 2016: 52-53). Guilt, resentment, pride, gratitude and admiration are thus examples of reactive attitudes that we naturally have when we are involved in inter-personal human relationships.[28]

[27] For further discussions of the connection between moral beliefs and motivation, see Björklund et al. (2012) and Kauppinen (2015).
[28] Some examples of reactive attitudes are amusement, reverence, envy, pity and embarrassment (Gibbard 1990: 135).

Furthermore, our moral beliefs that are related to our practices of blaming (and praising) can be thought to be constrained by the reactive attitudes. Following the example above, we know that we easily get blamed if we break a promise. Also, we naturally want to avoid being blamed by others because we want to avoid feeling guilty and also prevent others from resenting us (Gibbard 1990: 126). Hence, it could be argued that we come to have the moral belief that breaking a promise is wrong because of the prospect of avoiding those reactive attitudes (i.e., guilt and resentment) in the first place. If we have that moral belief, it is less likely that we will break a promise and thus get blamed for doing so in the future and, as a result, we would be less likely to feel guilty and get resented by others. As a result, this also means that having moral beliefs with the third feature also seems to be a constraint imposed by our reactive attitudes.[29]

(iii) Conclusion

In §2.4, I have suggested what kind of moral beliefs we have could be thought to be constrained by our emotional systems. According to the Exaptation Account, having moral beliefs with the three essential features are thus the constraint imposed by other adaptations (i.e., our emotional system) on our moral beliefs. If our moral beliefs with these features are really constrained by our emotional systems, it is then plausible to think that these beliefs are a by-product of our emotions. If this were true, our moral beliefs could be considered to be an exaptation.

In §2.3 and §2.4, I have outlined the Adaptation and the Exaptation Accounts in detail.

[29] Reactive attitudes are essential to us because they can help us coordinate our actions and expectations in social life (Gibbard 1990: 135). Hence, we evolved to have the reactive attitudes because they served a function for us in the circumstances where we initially participated in inter-personal social life. For a further discussion of reactive attitudes, see Strawson (1962/2008: 1-28).

I will remain neutral between which one of these accounts is true. Instead, as I will argue in the next Chapter 3 that, from the perspective of defending the EDA, it does not matter which one of these accounts is true as long as one of them is.

2.5 3th Theory of Explanation: The Cultural Evolution Account

In the previous sub-sections, I introduced two views of how evolution could have influenced our moral beliefs according to the biological evolutionary theories. However, many anthropologists, philosophers and sociologists have worked on a very different kind of evolutionary theory in recent years. Unlike the biological evolutionary theories, the new theories do not aim at explaining how biological traits and organs of organisms evolve. Instead, they are about how cultures evolve. Thus, this kind of evolutionary theories can be characterized as cultural evolutionary theories. Moreover, according to many cultural evolutionists, morality can also be explained in terms of cultural evolution, at least to a certain extent. They also claim that many essential features of moral beliefs may be explained in terms of cultural evolution. I will call this view the Cultural Evolution Account.

This sub-section consists of three parts. Firstly, I will explain what cultural evolution is, with an emphasis on the difference between it and biological evolution (2.5.1). Secondly, I will briefly introduce two dominant cultural evolutionary theories – the memetics account and the formal account (2.5.2). Finally, however, I will argue that the Cultural Evolution Account should be considered to be irrelevant in all discussions of

the EDA (2.5.3).

2.5.1 What Cultural Evolution is

In this sub-section, I will present a rough outline of the idea of cultural evolution. But before doing so, I first need to explain how I understand cultures. Although we are quite familiar with the idea of culture in our daily life, it is very difficult to give a unified definition of what the term really means. The difficulty also partly stems from the fact that culture consists of so many different things, including the likes of language, technology, art, literature, thoughts, "social and political institutions, ways of doing things" and even "in general all of the creations of the human mind" (Ayala 2010: 9015).

Even so, I think Eva Jablonka and Marion Lamb (2005) provide a brief but very insightful outline of the term 'culture'. They define culture as "a system of socially transmitted patterns of behaviour [including skills, practices, habits, beliefs], preferences, and products of animal activities that characterize a group of social animals" (Jablonka & Lamb 2005: 160). By defining culture in this way, Jablonka and Lamb highlight another very important feature of culture: Culture is not merely about the social patterns of different behaviour, preferences, ideas, and so on, but rather the idea of a culture is also about how these patterns are transmitted. How do cultural patterns change over time and across generations? How does culture transmit through different members in a society, as well as from one society to another society? Theories of cultural evolution thus aim at explaining the ways in which cultures change, transmit and persist in populations.

Cultures evolve and change in certain specific ways. This basic idea is shared by all theories of cultural evolution. As evolutionary theories, theories of cultural evolution also aim at explaining cultural evolution in terms of the three steps of natural selection, that is, *variation*, *reproduction* and *selection*. However, cultural evolution theories understand these three steps in a very different way than biological evolutionary theories.

Let us first consider the first step of natural selection: *variation*. It is obvious and evident that there is cultural variation. Contemporary and medieval Western societies have different living habits, knowledge, technology, and so on. Likewise, today, Western societies and East Asian societies also have different habits, patterns of behaviour, language, and so forth. Cultural evolutionists suggest that these cultural variations exist because the members of different societies have invented different cultures in their own societies and populations.

Next, we can consider *reproduction*, which is the second step of natural selection. Cultures can be inherited and thereby transmitted in a population. However, the process of reproduction is entirely different in this case when compared to sexual reproduction in the context of biological evolution. Metaphorically, traits are transmitted vertically in biological evolution. In other words, parental traits are transmitted to and inherited by offspring via sexual inheritance across generations (Lewens 2015: 10). Yet, parents do not transmit any cultural traits to their offspring via sexual inheritance across generations. Instead, cultures are transmitted and inherited in asexual ways, including via the processes of social learning, teaching, imitation and so on (Boyd & Richerson

2005: 3). We usually acquire habits and patterns of behaviour from our parents. However, it is also possible and very common that we acquire different habits and patterns of behaviour from our peers, our teachers and other individuals around us. Metaphorically speaking, cultures reproduce and transmit in ways that are horizontal and oblique.

Finally, cultural evolution also involves *selection*, which is the third step of natural selection. Cultures, and their elements, can be transmitted from one society to other societies. Some cultures die out. Some cultures are widespread. This shows that there is a selective process here. But how does selection work in cultural evolution? We can understand selection in cultural evolution in terms of accumulation. That is, a cultural trait that is selected by a population when it is commonly and widely shared within the population (Boyd & Richerson 2005: 44).

Scientific knowledge and different new technologies are prime examples of how cultural traits are accumulated. Let's imagine that members of a population acquire a piece of new scientific knowledge, say, the Newtonian theory of classical mechanics. Now let us assume that this knowledge is accumulated. In other words, this means that this knowledge is inherited in the population, and based on this cultural trait, a population can acquire new traits, for example, celestial mechanics, new engineering skills, and so on. This shows that, if cultural traits could not be accumulated, it is unlikely that we humans could have come to have the scientific knowledge and technologies that we have today (Boyd & Richerson 2005: 44).[30]

[30] There is a debate on whether cultural evolution also occurs in non-human animals (Gruber et al. 2019;

There are many cultural traits that are accumulated in our societies and that are transmitted to the new members of the population through teaching and learning, for example, different scientific theories, tools making skills, agriculture practices, and so on. These cultural traits are the selected cultural traits in our societies. In comparison, there were many cultural traits that have died out, for instance, endocannibalism, duelling, wanton killings, and so on. This shows that there must be some kind of a selection process also for cultures and cultural traits. Some traits are better transmitted and adapted while some eventually die out.

This section has briefly explained the basic general crux of cultural evolution. But what exactly are the specific ways in which cultures actually evolve? Different theories of cultural evolution provide different answers to this question.

2.5.2 Two Dominant Theories of Cultural Evolution – A Brief Explanation

In this section 2.5.2, I will briefly introduce two dominant theories of cultural evolution – the memetics theory and the formal models theory. I will remain neutral on how plausible these theories are, as it is beyond the scope of this thesis to discuss that

Laland & Janik 2006). For a discussion of how cultural evolution seems to occur in Japanese macaque monkeys on a small island Koshima, see Jablonka and Lamb (2005: 178-179). I believe that cultural evolution also occurs in non-human animals. However, as Boyd and Richerson (2005) suggest, human social learning is very unique because other animals may not be able to "accumulate adaptive information over many generations, building complex artifacts and institutions composed of many small innovations" (16). As a result, even if cultural evolution could occur in non-human animals, it would be much less complicated than human cultural evolution. Unfortunately, it is beyond the scope of this thesis to further discuss the difference between culture evolutions in human and non-human animals. For further discussions on this issue, see Boyd and Richerson (1996) and Mesoudi and Thornton (2018).

question. But what I want to do here is to demonstrate how cultural evolution is possible according to these two very different theories.

(i) Memetics Theory

The memetics theory was first introduced by Richard Dawkins (1976) and then also defended and developed by Susan Blackmore (1999). As an evolutionary theory, the memetics theory also needs to posit something like the three steps of natural selection.

One of the most essential features of the memetics theory is that it takes biological evolutionary theory (or precisely speaking, the gene-driven biological evolutionary theory) as a serious model (Lewens 2015: 12). According to the gene-driven biological evolutionary theory, genes determine how biological traits develop in an organism. For example, the genes that human beings have are essentially different from the genes that fishes. As the genes decide how biological traits develop, lungs are developed in individual human beings while gills are developed in fishes.

By taking the gene-driven biological evolutionary theory as a serious model, the memetics theory assumes that there are some gene-like, information-containing replicators in culture (Jablonka & Lamb 2005: 207). These gene-like "cultural entities" are thought to be memes (Dennett 2001: 308). For example, fashion, lifestyles, music and ways of building arches are some notable examples of memes (Lewens 2015: 26). Memes are also responsible for the development of cultural variants in the same way as genes are responsible for the development of biological variants. According to the memetics theory, memes determine how different forms of culture are reproduced in

populations. For example, members of different societies may have different memes concerning music. As a result, popular musical cultures often end up being very different across different societies. As the memetics theorists would suggest, this is because different cultures of music are the products of different musical memes (Jablonka & Lamb 2005: 207).

Let us then focus on the second step of natural selection, that is, *reproduction*. According to Richard Dawkins (1982), meme is a "unit of cultural inheritance" analogically similar to gene (290). Hence, memes reproduce in a similar way in which genes reproduce. According to the gene-driven evolutionary theory, genes make copies of themselves, and these copies are transmitted from the parents to the offspring. Therefore, genes are thought to be a kind of replicators. Biological traits are transmitted and inherited from parents to the offspring by coping the genes from the former to the latter.

The memetics theory then takes the gene-like memes to be replicators that are responsible for cultural inheritance. Just like how genes replicate themselves, the memetics theory claims that memes too can copy themselves, and these copies are then transmitted from one brain to another. As a meme is "an information-packet with attitude" by nature, memes can be transmitted from one brain to another through social learning and imitation (Dennett 2001: 309). As a result, cultural traits are transmitted and inherited from one to others by coping the memes from the former to the latter.

Finally, let us consider the third step of natural selection – *selection*. At this point,

however, there is an essential difference between the gene-driven biological evolutionary theory and the memetics theory. In the gene-driven biological evolutionary theory, the fitness of relevant genes and the fitness of the species that carry these genes are closely related. If the species carrying the given genes died out, the genes would die out too. However, according to the memetics theory, the fitness of a meme can be irrelevant to the fitness of the host that carries the copy of the meme. It is possible that a meme can "selfishly" flourish in a population and at the same time undermine the net fitness of the hosts that carry a copy of the meme (Jablonka & Lamb 2005: 207).

For example, let us consider the meme that is responsible for smoking. We can also assume that the trait of smoking has a net negative effect on the fitness of the hosts that have a copy of this meme. However, suppose that this meme S "has strong socially contagious and addictive effects" that can copy themselves quite successfully in a population (Jablonka & Lamb 2005: 207). As a result, the meme that is responsible for smoking could flourish in the population, but the net fitness of the carriers of this meme is diminished at the same time.

In fact, the proponents of the memetics theory argue that in many cases the fitness of memes and the fitness their hosts coincide. For example, the memes that are responsible for eating less fatty food and sugar could arguably enhance the fitness of human beings if these memes flourished in our societies. However, memes by nature only enhance their own fitness.[31] Therefore, it is a famous slogan that memes are "selfish replicators"

[31] Metaphorically, the memetics theory usually compares memes to be something like viruses or parasites (Dennett 2001: 309). There are three possible relationships between the fitness of the host and its parasite: (1) The presence of the parasite decreases the fitness of the host; (2) the presence of the parasite is neutral

(Jablonka & Lamb 2005: 207; Blackmore 1999: 30).[32] If cultural evolution really works in accordance with the memetics theory, then we must also accept the following view: Not all cultures that flourish in a population enhance the net fitness of the population.

(ii) The Formal Models Theory

Another dominant theory of cultural evolution is the 'formal models theory', which is defended Robert Boyd, Joseph Henrich and Peter J. Richerson (Boyd & Richerson 1985, 2005; Henrich & Boyd 2002). To introduce this theory, let us begin by asking why different societies have different cultures. Unlike what the memetics theory suggests, the formal models theory argues that the "cultural transmission processes are usually incomplete and imperfect" (Henrich & Boyd 2002: 88). This is because the methods of cultural transmission such as social learning and imitation are prone to error. Elements of culture are not simply copied and replicated like the genes.

Consider a case in which I want to learn a new language as an example. It seems that the ability to speak a new language is simply not replicated and then transmitted from my teacher to me. After all, I will make many errors when I am learning the language. Hence, the cultural traits seem to be transmitted in a way that involves of many error-

with respect to the fitness of the host; or (3) the presence of the parasite increases the fitness of the host. According to Dennett (2001), it seems that there are also three different kinds of relationship between the memes and their carriers, which are very similar to the three kinds of relationships between the host and its parasite described above (309).

[32] There are many objections to the memetics theory. To name a few: There is a question of whether the analogy between genes and culture is mistaken (Henrich et al. 2008). Some people also suggest that cultural units are not replicators (Sperber 2000; Sterelny 2006; Jablonka & Lamb 2005: 207), or that cultures cannot be understood as discrete units like memes (Kuper 2000: 180; Jablonka & Lamb 2005: 210-212).

prone human psychological and cognitive processes. For instance, in the previous example, I need to memorize many new terms and their pronunciation. I also need to try to use them in conversation in order to know how to use them in the appropriate way. The formal models theory thus argues that it is common for there to be cultural variations because of the involvement of error-prone human psychological and cognitive processes when we acquire new cultural traits.

Let us then focus on how the formal models theory explains the transmission of culture without taking the biological evolutionary theory as a serious analogy. As Boyd and Henrich (2002) suggest, culture is heavily shaped by both "psychological processes that determine how people think and feel, and social processes that determine how people interact" (88). This means that the key to explain cultural transmission is to understand how the psychological and social processes can help an individual to acquire cultural traits.

As mentioned above, the ability to acquire cultural traits is error-prone. However, the formal models theory argues that it is still possible that there is a reliable connection between the change of social patterns and our individuals cognitive and psychological processes. And, the proponents of the theory under investigation build formal models of social learning to explain why there is a reliable connection between cultural inheritance and our cognitive and psychological processes (Henrich & Boyd 2002: 110). The formal models theory suggests that there are two kinds of cognitive biases that can overcome the effects of error-prone learning (Lewens 2015: 127-128).

The first bias is prestige bias. That is, individuals tend to acquire cultural traits from the people who are in prestigious positions (Lewens 2015: 116). For example, we are more likely to acquire cultural traits from the likes of professors, millionaires, leaders of the societies, and so on. The second bias is conformist bias. Instead of randomly acquiring cultural traits from others, we also have a cognitive bias to acquire cultural traits that are more common in the population. It is a very common thought that one should follow how the locals act in a new place. For example, when a westerner moves to an East Asian country, she may acquire the culture trait of bowing by following how the locals in this country act. As this trait is a gesture of respect in many East Asian countries, this westerner would be more respected by the locals if she greeted them by bowing.

It is obvious that these two cognitive biases are not infallible. For the prestige bias, it is possible that professors and millionaires have cultural traits that will fail to flourish. As for the conformist bias, common cultural traits are not necessarily always the traits that will flourish in the long run. Also, even if we acquired cultural traits through these two kinds of cognitive biases, this would not always guarantee reproductive success. Nevertheless, according to the formal models theory, these two cognitive biases (perhaps alongside other kinds of biases) still play a significant selective role in cultural transmission, and we can have a greater chance to flourish with these two cognitive biases.[33]

I would like to end this sub-section with a reminder. I will put aside the question of

[33] According to Morin (2016a), one of the biggest problems of this theory is how to understand the "interactions between culture and cognition" essentially in terms of prestige and conformist bias (447). For other challenges to the formal models theory, see Lewens (2015: 114-128) and Morin (2016b).

which theory of cultural evolution is the most plausible one, as it is beyond the scope of my thesis to discuss that question further. Nevertheless, I will assume that both theories should be taken into consideration when we attempt to explain how we came to have our moral beliefs. It may well be that many of these beliefs are a result of the previous types of processes of cultural evolution, or so I will argue next.

2.5.3 Explaining Our Moral beliefs in Terms of Cultural Evolution

In this sub-section, I will explain how our moral beliefs could be explained in terms of cultural evolution. In the previous sub-sections 2.3.2 and 2.4.2, I introduced two biological explanations of how we came to have our moral beliefs. Although these two explanations, as I have argued, are quite plausible, there are serious questions over whether morality can be fully explained by relying merely on biological evolutionary theory. Many cultural evolutionists have thus attempted to explain at least some elements of human morality by relying on cultural evolution.

However, at this point, it should be emphasized that our moral beliefs can be explained at two different levels. On the one hand, we can explain how we came to have the moral beliefs we have. On the other hand, we can also describe the exact mechanisms through which we came to form these moral beliefs. In this section 2.5.3, I will first illustrate how moral beliefs can be explained at two different levels. I will then argue that the cultural evolutionary theorists agree with the biological evolutionary theorists about why we came to have the moral beliefs we have. The only thing that they disagree about is the exact mechanisms through which we came to form these moral beliefs. For the purpose of the thesis (i.e., to make a defence of the EDA), it makes no real difference

that which mechanisms are the exact one through which we came to form these beliefs.

(i) The proximate and the ultimate causes

In order to illustrate how moral beliefs can be explained at two different levels, I need to introduce a well-known distinction from biology between proximate causes and ultimate causes. This distinction was first suggested by evolutionary biologist Ernst Mayr (1961), and it is supposed to distinguish between "two [different] types of causal explanation" that describe what happens in natural selection (Levy & Levy 2016: 2; Scott-Phillips et al. 2011: 44). Let us start with the explanations at the level of ultimate causes. The ultimate cause of a trait is whatever caused that trait to be selected in the natural selection. Therefore, an ultimate explanation of a trait focuses on the "fitness consequences" of the trait in order to give an account of why that trait was favoured by the natural selection. In contrast, a proximate cause of a trait is the physiological or a psychological mechanism that is responsible for the trait in question in the given species. Therefore, a proximate explanation of a trait concerns how individuals and members of the species come to inherit and develop the trait.

Let us consider bird migration as an example to illustrate the difference between an ultimate cause and a proximate cause (Mayr 1961: 1502-1503). There are some birds, such as warblers, which spend "different seasons in different locations" (Levy & Levy 2016: 2). For instance, warblers migrate to warmer areas every winter. We may then ask: What caused the warblers to have this trait? At this point, we need to recognize that we can answer this question at two different levels by either providing the ultimate causes or the proximate causes of the relevant behaviour. We can thus provide an

ultimate explanation and a proximate explanation by outlining the different causes of the same trait. These two explanations are, however, addressing different questions of the origins of this trait at different levels.

Let us begin with the ultimate explanation of the warblers' migration. An ultimate explanation of this trait needs to explain why it is better for warblers to have this trait in terms of what consequences the trait has for their fitness (Mayr 1961: 1502). For example, several ecological factors could be argued to be a central element of the ultimate cause of this trait, as there could be significant differences in food availability and climate in different locations (Levy & Levy 2016: 2; Mayr 1961: 1502). If the warblers didn't migrate to a warmer place in the winter, then most of them would probably starve to death (Mayr 1961: 1502). As a result, this type of ecological factors constitutes the ultimate cause of the relevant behaviour. This is because those ecological factors explain why this trait was favoured in natural selection.[34]

In contrast, the proximate explanation of the same trait of warbler migration is completely of a different kind. As already mentioned, the aim of the proximate explanation is to describe the mechanism through which individual warblers come to acquire this trait. Mayr (1961) suggests that there are certain immediate external factors (which include the weather conditions and the length of day or night) that affect the warblers (1503). And, there are also certain physiological mechanisms in the warblers

[34] A trait can have more than one ultimate cause. For example, in addition to the ecological cause, warblers seem to come to have the trait of migration also because of the genetic constitution. As Mayr (1961) suggests, "the warbler has acquired a genetic constitution… which induces it to respond appropriately to the proper stimuli from the environment" (1502). Without this genetic constitution, warblers would not be able to migrate in winter. Therefore, the genetic constitution can also be considered to be another ultimate cause of the trait of warbler migration.

themselves that react to these external factors. By responding to the "proper stimuli from the environment", the warblers will start to migrate and to actually take off (Mayr 1961: 1502-1503).[35] The proximate causes of the trait of warbler migration are thus those physiological mechanisms that react to the external factors and cause the migration.

(ii) The explanation of moral beliefs in virtue of cultural evolution

Let us then return to the case of moral beliefs. Just like many other traits, we can provide both ultimate and proximate explanations of our moral beliefs. However, these two explanations are explaining the very same sets of moral beliefs but just in two different ways at different levels. They are complementing explanations rather than mutually exclusive ones. The ultimate explanations of our moral beliefs concern why our moral beliefs with the three essential features were selected for in the natural selection, while the proximate explanation of the same moral beliefs concerns the mechanism through which we come to have these moral beliefs.

Let us start from considering the ultimate explanations of our moral beliefs. In the previous sections 2.3.2 and 2.4.2, I have explained how it could be argued that we came to have our moral beliefs because such beliefs were either an adaptation or an exaptation. The explanations outlined in those sub-sections were at the level of the ultimate explanations of our moral beliefs. This is because both explanations concern the question of why our moral beliefs were selected for in natural selection.

[35] For example, warblers will start to migrate "as soon as the number of hours of daylight have dropped below a certain level" and also migrate following a particular route because of the wind and temperature conditions (Mayr 1961: 1503).

In 2.3.2, I already explained why having moral beliefs with core contents could have helped our ancestors to co-operate and to avoid the escalation of violence and thus ultimately the collapse of their societies. As a consequence, if our moral beliefs were an adaptation, then having moral beliefs should have helped our ancestors to leave more offspring in the succeeding generations. This is why it could be argued that our moral beliefs were also selected for in natural selection and thus the explanation of why we came to have moral beliefs in virtue of adaptation is also an ultimate explanation of our moral beliefs.

If the evolutionary debunkers instead think that our moral beliefs are an exaptation, then they are in that case committed at least to the claim that our moral beliefs are currently adaptive. This is because, according to the definition of exaptation, it is necessary that an exaptation is currently adaptive for a certain function. Hence, if our moral beliefs are thought to be an exaptation, then they should also be thought to be currently adaptive and thus to have a certain function that effectively enables human beings to leave more offspring. Hence, if evolutionary debunkers want to argue that our moral beliefs are an exaptation, then they are also required to explain why our moral beliefs are currently adaptive for a certain function. As a result, this explanation of why our moral beliefs are currently adaptive is also kind of ultimate explanations of our moral beliefs.

In contrast, the explanation of our moral beliefs in terms of cultural evolution should not be considered to be the same kind of an ultimate explanation of our moral beliefs

as the previous ones. If the cultural evolutionary theorists really want to provide an ultimate explanation of our moral beliefs, they are required to explain why our moral beliefs, as a cultural trait, would have been selected for in cultural selection. In 2.5.1, I mentioned that theories of cultural evolution merely aim at explaining the ways in which cultures change, transmit and persist in populations. Hence, even if our moral beliefs were really a cultural trait, cultural evolutionary theorists can only explain how moral beliefs are culturally transmitted and why a moral belief, as a cultural trait, can thus be widespread in a community. As a result, the kind of explanations of our moral beliefs based on cultural evolution does not really concern the question of why our moral beliefs were selected for in natural selection and therefore should not be considered to be an ultimate explanation of our moral beliefs.

Although cultural evolutionary theorists are unable to provide an ultimate explanation of our moral beliefs, they can still explain the proximate mechanisms of our moral beliefs. A plausible proximate explanation of our moral beliefs aims at describing the mechanism that is causally responsible for producing moral beliefs in us, and the explanation of our moral beliefs based on the idea of cultural evolution should be considered to be a kind of proximate explanations. Moreover, the biological and cultural evolutionary theory could also be thought to be competing theories of what the proximate mechanism is. Those theories disagree with the cultural evolutionary theories in this domain about whether our moral beliefs are transmitted culturally or inherited biologically.

Which of the two views of the transmission of our moral beliefs is then more plausible?

As explained in the previous sub-sections 2.5.1 and 2.5.2, cultures are transmitted via processes such as social learning and imitation. If having moral beliefs is a cultural trait, we should be able to acquire new kinds of moral beliefs much more quickly (for example, because of how quickly memes can spread). In contrast, if our moral beliefs were biologically inherited (via genetic processes and sexual reproduction), then the changes of our moral beliefs would take place much more gradually.

I believe that the cultural view of the transmission of our moral beliefs is more likely to be true. For example, it is well-documented that Eskimos would kill their newborns in the past. Assume that it was a moral belief of Eskimos that it is right to kill their newborns. Without enough resource, it could be reproductively advantageous for Eskimos to kill some newborns. However, it is very difficult to explain how they could inherit this belief biologically. In contrast, it is much easier to explain how Eskimos could acquire this moral belief culturally because cultures have been transmitted in quick ways, including via the processes of social learning, imitation and so on. Then, it is very possible that all Eskimos could acquire that moral belief from their peers around them.

Nevertheless, in this thesis and in all the discussions of the EDA, what we really need to focus on is the ultimate explanations of our moral beliefs instead of their proximate explanations. This is because, as I will explain in Chapter 3, evolutionary debunkers aim to argue that the ultimate causes of our moral beliefs are generally such that they are not tracking the moral truth, but they do not aim at describing the exact mechanisms through which we came to form these moral beliefs. Although the biological and

cultural evolutionary theories disagree on what the proximate mechanisms are, these mechanisms are really irrelevant to whether we can provide plausible ultimate explanations of our moral beliefs such these beliefs end up not tracking the moral truth.[36] Hence, in this thesis, we can put aside whether our moral beliefs can be better explained in terms of cultural evolution or biological evolution.

2.6 Conclusion

In this chapter, I have introduced and evaluated three potential explanations of how evolution could have influenced and shaped our moral beliefs. Based on what I have argued in this chapter, I suggest that both the Adaptation and the Exaptation Accounts have their own plausibility: both of them offer plausible views of how evolution could have influenced and shaped our moral beliefs. In this thesis, I will remain neutral between whether Adaptation Account or the Exaptation Account is true. This is because, as I will argue in the next Chapter 3, our moral beliefs can be thought to be not tracking the moral truth as long as one of these accounts is true. Hence, from the perspective of defending the EDA, it does not matter which one of them is true.

Finally, in the end of this chapter, I suggested that evolutionary debunkers could also have a disagreement about whether our moral beliefs are transmitted culturally or

[36] There is no reason why the biological and cultural evolutionary theorists could not agree on the same ultimate cause of our moral beliefs. After all, they just disagree with each other regarding the proximate causes of our moral beliefs. The proponents of the cultural evolution theories can simply accept the ultimate explanations of our moral beliefs provided either by the defenders of the Adaptation Account and the Exaptation Account. For instance, they too can say that the moral beliefs with core contents were selected for in natural selection because the core contents of these beliefs enabled co-operation and prevented violence and conflicts (and as a result, having moral beliefs with these core contents enable us to leave more offspring).

inherited biologically. Nevertheless, I also argued that such a disagreement is actually irrelevant to whether the evolutionary debunkers are able to provide plausible ultimate explanations of our moral beliefs such these beliefs end up not tracking the moral truth.

Chapter 3

Evolutionary Origins of Moral Beliefs and Truth-Tracking

3.1 Introduction

In the previous Chapter 2, I explained how evolution has influenced and shaped our moral beliefs. I argued that both Adaptation Account and the Exaptation Account can offer plausible views of how evolution could have influenced and shaped our moral beliefs. Nevertheless, even if either one of these accounts were true, this does not necessarily imply that our moral beliefs would not be tracking the moral truth reliably, as argued by the evolutionary debunkers (Kahane 2011: 115; Vavova 2015: 107). As a result, the evolutionary debunkers are required to explain why our moral beliefs would not be tracking the moral truth reliably due to their evolutionary origin.

Therefore, this chapter 3 will address an important question that needs to be answered by both the evolutionary debunkers and also their opponents: If evolution has shaped and influenced our moral beliefs in the ways explained in the previous Chapter 2, does this also mean that our moral beliefs are not tracking the moral truth reliably? In this Chapter 3, I will conclude that the answer to the previous question is 'Yes': Our moral beliefs are not tracking the moral truth reliably due to the evolutionary origins of these

beliefs.[37]

But before we get to the previous question, we need to first consider the notion of truth-trackingness itself. At first glance, it seems like the whole idea of what it is to track the truth is not very clear at all (Bedke 2014: 105). This is because this notion is nothing more than a metaphor suggesting a certain kind of a relation between the relevant truths and our beliefs. Roughly speaking, if our beliefs are tracking the truth, then these beliefs are formed in a way in which they "adequately align" with the corresponding truths (Bedke 2014: 108; Kahane 2011: 105). There are, however, more than one possible way in which our beliefs could be aligning with the corresponding truths. Hence, the metaphor of truth-trackingness itself tells us nothing precisely about what this alignment actually is.

There are two dominant views of how to understand of what kind of alignment truth-trackingness consists – (i) the explanatory reading of truth-trackingness (Joyce 2016b: 132-137; Street 2006: 160n35) and (ii) the modal reading of truth-trackingness (Joyce 2016b: 129-132; Korman & Locke forthcoming; Vavova 2014: 90n27). In accordance with these two different readings of truth-trackingness, there are two different ways in which evolutionary debunkers could argue that our moral beliefs do not track the moral truth.

This chapter thus consists of three parts. Firstly, §3.2 will introduce the Explanatory

[37] In this chapter, I will put aside the question of why the epistemic status of our moral beliefs would be undermined if they were not tracking the moral truth. I will address this question in Chapter 4.

Account of truck-trackingness and its two variants – the causal version and the liberal version. In this §3.2, I will also try to reject the causal version and explain why, according to the liberal version, our moral beliefs should be seen as not tracking the moral truth due to their evolutionary origin. Secondly, in §3.3, I will introduce the Modal Account of truck-trackingness and consequently also of why, according to that view too, our moral beliefs should be seen as not tracking the moral truth due to the evolutionary origins of these beliefs.

Finally, in §3.4, I will introduce a potential objection to the evolutionary debunkers which is suggested by Justin Clarke-Doane (2012; 2014; 2015). According to Clarke-Doane, our moral beliefs could be argued to fail to track the moral truth only if we understood the notion of truth-trackingness by relying on the explanatory reading that is not to be very plausible more generally. He then argues that, if we understand the notion of truth-trackingness on the basis of the modal reading, then our moral beliefs could be argued to track the moral truth.

In §3.4, however, I will also argue that Clarke-Doane's objection fails for two reasons. Firstly, despite Clarke-Doane's concerns, evolutionary debunkers can still use the explanatory reading of truth-trackingness to argue our moral beliefs are not tracking the moral truth reliably due to the evolutionary origin of these beliefs. Secondly, evolutionary debunkers can also accept a version of the so-called anti-luck epistemology that is based on a revised safety condition. As a result, most, if not all, of our moral beliefs turn out not be safe in the required way and thus do not track the moral truth, or so I will argue at the end of this chapter.

3.2 1st Theory of Truth-Tracking: The Explanatory Account

I will start from how the notion of truth-trackingness has been understood by the Explanatory Account in the case of ordinary non-moral beliefs. According to the explanatory reading of truth-trackingness, a given set of beliefs in a particular domain tracks the truths of that domain – including its facts and properties – if there is an explanatory connection between those beliefs and the relevant truths (Korman & Locke forthcoming). In order to make sense of this account, we then need to understand what kind of an explanatory connection would be required.

Consider some of our empirical beliefs of water as an example. For instance, we believe that water is colourless, liquid at room temperature, odourless and so on. How did we all come to have these beliefs? Arguably, we came to have these beliefs on the basis of perceiving things by using our vision and sense of smell. Yet, some sceptics could argue in this situation that it is still possible that we all are being deceived by an evil demon about the previous things. For example, the sceptics could suggest that maybe we came to have the previous water-beliefs merely because a devil wanted us to have those beliefs. This suggests that there is more than one possible explanation of how we came to have our water-beliefs.

According to the explanatory reading of truth-trackingness, however, the best explanation of how we came to have our beliefs about water (for example, that it is colourless, odourless, and so on) is that there indeed exists water that has the properties of being colourless, odourless, and so on (Street 2006: 160n35). This is not to deny that

other potential explanations of how we came to have these beliefs about water do not exist. However, it is most likely that we came to have these beliefs because the facts concerning water, in reality, play an explanatory role: They are a part of the process through which we came to have our water-beliefs (Sayre-McCord 1988: 262).[38]

Hence, according to the Explanatory Account, the previous beliefs about water can be argued to be tracking water truths precisely for the previous reason. This example can then be used to illustrate the suggestion that there is thus a condition that truth-tracking beliefs must meet: A belief about X tracks the relevant truths about X if and only if the best explanation of how this belief about X was acquired must rely on the truths about X itself (Street 2006: 160n35). Our beliefs are tracking the relevant truths if and only if this condition is satisfied.

Nevertheless, there are two different views of what counts as the best explanation of a belief that relies on a given set of facts. As a result, there are also two different versions of explanatory reading of truth-trackingness in accordance with these two views. I will first introduce and evaluate the causal version in the following sub-section 3.2.1 and the liberal version in the following sub-section 3.2.2 respectively. I will also argue why the causal version cannot be used for the purposes of the EDA.

[38] Some sceptics may argue that we should not assume the existence of water in order to explain how we came to have the beliefs about water. However, as Nicholas Sturgeon (1988) suggests, the reference to the relevant fact is needed in the explanation of a belief if that explanation "is somehow better than competing explanations [of those beliefs]" (236). Hence, in this case of our beliefs about water, it is reasonable for us to assume the existence of water because the explanation of our beliefs about water that refers to facts about water is better than the other explanations of those beliefs (Harman 1988: 121-122).

3.2.1 The Causal Version of Explanatory Reading of Truth-trackingness

Let us begin with the causal version of the Explanatory Account first (Artiga 2015: 3361; Fraser 2014: 458-461; Kahane 2011: 106). As we just saw, the explanatory reading of truth-trackingness claims that a belief about X tracks the relevant truths about X if and only if the best explanation of how this belief about X was acquired must rely on the truths about X. According to the causal version of Explanatory Account, an explanation of how a belief about X was acquired can count as the best explanation only if this explanation is a causal one. Therefore, according to the resulting causal version, the only kinds of explanations how acquired the relevant beliefs that can satisfy the condition set for truth-trackingness are causal ones, and so according to the resulting view a belief about X tracks the relevant truths about X if and only if the truths about X at least in part cause the relevant believer to have this belief about X.

Let's consider the pen on my desk as an example. Let us also assume that it is a fact that there really is a pen on my desk. As an object, the pen has causal powers that can lead us to form certain beliefs (Street 2016: 322). With its causal powers, the pen can cause me to come to have the belief that there is a pen on the desk through my sense of sight. When I see the pen on the desk, the pen thus causes me to form the belief that there is a pen on my desk. The causal explanation of how I came to have this belief that there is a pen on my desk must thus refer to the relevant facts about the pen. Therefore, my previous belief also satisfies the relevant condition for being a truth-tracking belief: It tracks the relevant truths about the pen on the desk precisely because the best explanation of how this belief was acquired must rely on the relevant causal connection

between the pen and the belief.

We can then return to the case of moral beliefs. According to the causal version of the Explanatory Account, for our moral beliefs to track the relevant moral truths, they too would have to satisfy the following causal condition of truth-trackingness: Our moral beliefs track the moral truth if and only if the moral truth at least in part causes us to have our moral beliefs. However, it is problematic for the evolutionary debunkers to use this causal notion of truth-trackingness in the EDA, or so I will argue in the rest of this sub-section 3.2.1.

Let us start from considering one significant feature of the EDA, namely that the EDA is a targeted argument. Its target is to show that certain views in meta-ethics fail for epistemic reasons. At least in principle, the EDA could be understood to pose a threat to all meta-ethical views according to which there are moral truths as well as justified true moral beliefs about them and thus moral knowledge.[39] More specifically, moral realism is widely accepted to be one of the most significant targets of the EDA. For example, Sharon Street (2006) suggests that the target of her version of the EDA is the "realist theories of value" (110). Some versions of the EDA may be refined to have an even narrower target. For example, Katia Vavova (2015) argues that the EDA should target the view holding that "moral truths are attitude-independent", or in other words,

[39] The main reason to emphasize that the EDA is a targeted argument is to distinguish it from more global sceptical arguments. Some philosophers have investigated whether the EDA would debunk too much (Vavova 2014; Crow 2016). More precisely, they have focused on whether we could debunk all kinds of knowledge – including knowledge of the external world and the scientific knowledge – based on the same line of reasoning as the one used in the EDA. For objections to the claim that EDA would debunk too much, see Bedke (2009: 203-204), Moon (2017), and Street (2006: 159n25).

moral realism (108).[40] In my thesis, I will only focus on the forms of the EDA that target moral realism.[41]

Let us then consider the significant distinction between the naturalist and non-naturalist versions of moral realism (Finlay 2007: 826-829; Fisher 2011: 81). Both versions of moral realism are often taken to be the natural targets of the EDA. Nevertheless, it is important to emphasize that there is a significant difference between these versions of moral realism. Because of this difference, the causal reading of truth-trackingness should not be used in the EDA for a certain reason.

According to the naturalist versions of moral realism, moral facts and moral properties are natural. Furthermore, there are at least three different views of what it is for moral facts and properties to be natural facts and properties. Briefly, moral facts and moral properties have been thought to be natural because they can be (i) be reduced to natural properties by doing an *a priori* conceptual analysis of our moral terms (Finlay 2007: 828; Fisher 2011: 69);[42] or (ii) be reduced to natural properties via an *a posteriori* investigation (Finlay 2007: 840; Fisher 2011: 70);[43] or they (iii) are open to an

[40] This view is defended by ethical naturalists (including Boyd (2007) and Jackson (1998)) and ethical non-naturalists (including Cuneo (2007), Enoch (2011), and Shafer-Landau (2003)). For overviews of moral realism, see Finlay (2007) and Fisher (2011: 55-90).

[41] This is not to say that all other meta-ethical views are immune to the EDA. Nevertheless, it is hard to deny that moral realism is the most significant and common target of the EDA. Moreover, if the EDA failed to target moral realism, it would become a much less interesting and meaningful argument. This is why the evolutionary debunkers cannot set moral realism aside when they try to construct the EDA. For representative defenders and opponents of the evolutionary debunking arguments who explicitly see moral realism as the primary target of the EDA, see Bogardus (2016), Das (2016), FitzPatrick (2014; 2015), Shafer-Landau (2012), and Street (2006). For an objection to the view that it is necessary for the EDA to target moral realism, see Sinclair (2018: 98-100).

[42] This view is also well-known as the "Canberra Plan". See Braddon-Mitchell and Nola (2009) and Jackson (1998). Also, I will further discuss this view in Chapter 5.

[43] See Railton (1986: 2003) and Schroeder (2007: 61-83).

empirical and *a posteriori* investigation even if the moral facts and properties are not reducible to any other descriptive properties – including physical, biological, psychological properties, and so on. (Zhong 2010: vi).[44]

The non-naturalist moral realists, in contrast, argue that moral facts and properties are not natural facts and properties. They claim that these facts and properties cannot be reduced to, identified with or realized by natural facts and properties. According to these non-naturalist moral realists, this also means that there exists of moral properties that are "*sui generis*" and "discontinuous" with other kinds of properties, including the natural properties (Väyrynen 2018: 171; Wielenberg 2014: 14). Hence, on this view, since moral properties are ontologically distinct or separate from the natural properties, they should also be considered to be causally inefficacious and inert (Klenk 2017: 795n11; Lutz 2018: 1106). In other words, moral facts and moral properties have, according to the non-naturalists, no independent causal powers (Lutz 2018: 1106; Shafer-Landau 2012: 217; Street 2016: 322).[45]

We can now return to the idea that the EDA is a targeted argument. The evolutionary debunkers mainly want to provide an argument against non-naturalist realism, which is one of the most significant targets of the EDA. Yet, if those debunkers also accept the causal Explanatory Account of the truth-trackingness, then they would not even need the EDA in the first place – the argument would become wholly redundant. This is

[44] This view can be called "non-reductive naturalist moral realism". See Sturgeon (1988; 2006: 96-99).
[45] It is beyond the scope of this chapter to explain further the distinction between the naturalist and non-naturalist forms of moral realism. For extended discussions of this distinction, see Finlay (2007: 826-829) and Shafer-Landau (2003: 58-65).

because, in this case, the debunkers can directly argue that our moral beliefs would trivially not be truth-tracking the alleged non-natural moral facts, regardless of the evolutionary origins of these beliefs.

Above I explained that, if the evolutionary debunkers accept the causal version of the Explanatory Account, then according to them our moral beliefs would track the moral truth if and only if the relevant moral truths would at least in part cause us to have our moral beliefs. However, as previously mentioned, the non-naturalist moral realists hold the view that moral facts and properties are causally inefficacious and inert, by definition. If moral facts and properties had no causal powers in this way, then these facts and properties could not causally explain our moral beliefs as well. This also means that it would be impossible for the moral facts and properties to be in even a part of what caused us to have our moral beliefs. As a result, our moral beliefs would fail to track the moral truth, regardless of the evolutionary origins of these beliefs. The causal version should thus not be used by the evolutionary debunkers, since the EDA itself would become redundant as an objection to the non-naturalist moral realists, the intended target of the argument.

Unlike the non-naturalist moral realists, however, the naturalist moral realists argue that moral properties are causally efficacious and do have certain causal powers (Klenk 2017: 795n11). This is because, on their view, moral properties are the same kinds of properties as other ordinary natural properties, which are causally efficacious. Thus, according to naturalist metaethical views, it would be at least possible for the moral facts and properties to be a part of what caused us to have our moral beliefs. Even if the

evolutionary debunkers accepted the causal version of the Explanatory Account, it would thus not be the case that our moral beliefs would be trivially not truth-tracking the alleged natural moral truths.

Yet, I believe that the causal version of the Explanatory Account should not be used in the EDA against the naturalist versions of moral realism for a completely different reason: The causal version is itself a bad account of truth-trackingness and therefore it should not be used in the EDA against any view. To illustrate this, let us start from the well-known example of fake barns, which was first suggested by Alvin Goldman (1976) in his article "Discrimination and Perceptual Knowledge" (772-773). Consider the following first case. Imagine that Harry is driving in the countryside where there are a lot of barn-like objects. At a certain point, he then looks at a particular barn-like object on a field next to the road. We can assume that Harry has excellent eyesight and has enough time to look at the object (Goldman 1976: 772). We can also assume that there are no obstacles between Harry and the barn-like object and the weather is fine as well. Harry then looks at this object and believes that it is a barn.

Nonetheless, unknown to Harry, the area is actually full of fake barns (Goldman 1976: 773). Let us also assume that these fake barns are "cleverly constructed" and look exactly the same as the real barns from the road (Goldman 1976: 773). These fake barns, however, just have barn-like facades. They have no back walls and thus cannot be used as real barns. As it happens, the barn-like object on the field that Harry is currently looking at is a genuine barn. Yet, if he encountered a fake barn, he would also take it to be a barn (Goldman 1976: 773).

In this situation, because Harry sees the object that is a real barn on that field, this object causes him to come to have the belief that there is a barn right next to him (Goldman 1976: 773). Even though there are lots of fake barns in the area, this fact has no influence on the causal explanation of how Harry comes to have the previous belief. The causal explanation of Harry's belief remains unchanged, and therefore the truth-tracking condition is still satisfied. As a result, according to the causal Explanatory Account of truth-tracking, Harry's belief would still be truth-tracking.

Intuitively, however, it is very hard to accept that Harry's belief (that there is a barn next to him) would be truth-tracking in the previous example. Had Harry been next to a fake barn on the field (and there are many fake barns in the area), Harry would still come to have the same belief. Thus, we need to find a more plausible view of truth-tracking that would enable us to avoid the conclusion that Harry's belief in the previous case is truth-tracking.[46] Furthermore, the previous case also shows why the causal versions of the Explanatory Account are so implausible and thus should not be used in the EDA.

To sum up, the causal version of Explanatory Account of truth-tracking cannot be used for the purposes of the EDA. Firstly, it makes the EDA redundant when it is used against the non-naturalist moral realists. Secondly, it is itself not very plausible, as I have just

[46] By using this example, I merely focus on the question of whether Harry's belief in the previous case is truth-tracking, whereas Goldman and many other epistemologists are more interested in the question of whether Harry in that case has knowledge of the barn (Goldman 1976: 773). See Feldman (2003: 88-90).

argued. Because of these problems, I suggest that the evolutionary debunkers should move on to consider another version of the Explanatory Account, as I will do next too.

3.2.2 The Liberal Version of Explanatory Reading of Truth-trackingness

Let us then consider another version of the Explanatory Account, that is, the liberal version. As previously mentioned, the proponents of the Explanatory Account generally suggest that our beliefs are tracking the relevant truths when the following condition is satisfied: A belief about X tracks the relevant truths about X if and only if the best explanation of how this belief about X was acquired must rely on truths about X. As explained above, the proponents of the causal version argue that the best explanations of how we come to have the relevant beliefs must always be causal explanations. In order to avoid the previous problems with the causal version of the Explanatory Account, the defenders of the liberal version argue that the best explanation of how we come to have the relevant beliefs does not necessarily need to be a causal one (Bedke 2014: 105).

Therefore, according to the liberal version, the best explanation of how a given truth-tracking belief about X was acquired really must rely on the truths about X, but it can be either a causal or a non-causal explanation. This also means that causal explanations could be but need not always be the best explanations of our beliefs that satisfy the truth-trackingness condition of the Explanatory Account. Let us then consider how the evolutionary debunkers could use this liberal explanatory view of truth-trackingness to argue that our moral beliefs are not tracking the moral truth. The basic crux of their

argument can be formulated as follows:

The Evolutionary Off-Track Argument (Liberal Explanatory Version)

1. The origin of our moral beliefs can be wholly explained by natural selection (the conclusion of Chapter 2).

2. If the origin of our moral beliefs can be wholly explained by natural selection, then moral facts and moral properties do not play an explanatory role with regards to how we came to have our moral beliefs.

3. Moral facts and properties do not play an explanatory role with regards to how we came to have our moral beliefs (1, 2, *Modus Ponens*).

4. If moral facts and moral properties do not play an explanatory role with regards to how we came to have our moral beliefs, then those facts and properties are not part of the best explanation of our moral beliefs.

5. Moral facts and moral properties are not part of the best explanation of our moral beliefs (3, 4, *Modus Ponens*).

6. A belief about X is truth-tracking if and only if the relevant facts and properties about X are a part of the best explanation of how we come to have this belief about X (the Explanatory Account).

Conclusion: Our moral beliefs are not tracking the moral truth (5, 6)

This argument is clearly valid. In the rest of this sub-section 3.2.2, I will focus on explaining the premises 1-6 one after another. Let us begin from why the premise 1 should be thought to be true. In Chapter 2, I introduced two plausible explanations of how evolution could have influenced and shaped our moral beliefs – the Adaptation Account and the Exaptation Account. As I mentioned in that chapter, I remain neutral between which one of these accounts is true. This is because, from the perspective of defending the EDA, it does not matter which one of them is true as long as one of these accounts is true. If either one of them is true, then we are able to wholly explain the origin of our moral beliefs in evolutionary term (either in virtue of adaptations or exaptations). As a result, because of what I have argued in Chapter 2, the premise 1 can be argued to be true.

We can then consider the premise 2. Let us consider two completely independent facts p and q. If a belief about X can be w*holly explained* by the fact that p, then it is obvious that the other fact that q does not play any explanatory role with regards to how we come to have this belief X as long as p and q are independent of one another. Hence, the premise 2 seems to be trivially true.[47] And, if the premises 1 and 2 of the previous argument are both true, then the premise 3 will be true as well by *modus ponens*.

[47] A possible objection to premise 2 is to argue that moral facts and properties could still play an indirect explanatory role with regards to how we came to have our moral beliefs even if we are able to explain the origin of our moral beliefs in evolutionary terms. This objection is often taken to be the 'third factor objection' against the EDA. In Chapter 6, I will discuss this objection in detail.

Let us then consider the premises 4 and 5. Given that the premise 3 is true, the explanation of how we acquired our moral beliefs does not rely on the relevant moral truths, which therefore have no explanatory role with regards to how we came to have those beliefs. Intuitively, then, good explanations, let alone the best explanation, of how we acquired our moral beliefs do not rely on any moral truths. As a result, the moral truths, whatever they could be, are not part of the best explanation of our moral beliefs. Thus, the premise 4 is true. And, then the premise 5 will be true as well by *modus ponens* if the premises 3 and 4 of the previous argument are both true.

Finally, the premise 6 is the previously mentioned definition of the liberal Explanatory Account. According to that definition, our moral beliefs track the relevant moral truths if and only if the relevant moral facts and properties are a part of the best explanation of how we come to have our moral beliefs. Yet, if the premise 5 is true, the best explanation of our moral beliefs does not rely on any moral facts or moral properties, and therefore these moral facts and properties are not a part of the best explanation of our moral beliefs. As a result, our moral beliefs are not tracking the moral truth, and the conclusion that our moral beliefs fail to track the moral truths is thus true.

Non-naturalist moral realists – one of the most obvious targets of the EDA – may try to reject the previous argument by denying its premise 2.[48] According to the liberal

[48] This Chapter 3 will not consider further the question of whether the liberal view of truth-trackingness could be used in the EDA against naturalist moral realism. As mentioned in 3.2.1, according to naturalist moral realism, moral facts and properties are causally efficacious. Thus, it could be argued that those moral facts and properties can be used in good causal explanations of how we came to have our moral beliefs. Yet, it is still unclear how, on the naturalist realists' view, moral properties could be a part of the evolutionary explanation of the origin of our moral beliefs (given that I already argued that how evolutionary debunkers can wholly explain the origin of our moral beliefs in Chapter 2).

account, the best explanation of our moral beliefs can be either a causal or a non-causal one. Even if the non-naturalist moral realists cannot provide a causal explanation of how we came to have our moral beliefs (after all, they think that moral facts are causally inert), they could try to provide a good non-causal explanation of how we came to have these beliefs that refers to the relevant moral truths.

To illustrate this objection, let us consider an example suggested by Richard Joyce in his article "Reply: Confessions of a Modest Debunker" (2016b). Imagine that Anna comes to believe that 7 is a prime number (Joyce 2016b: 133). Numbers are generally considered to be causally inert, just like the moral properties according to the non-naturalist moral realists (Bell & Hart 1979: 163).[49] This means that the number 7 itself could not have caused Anna to have the previous belief and so it could not have played any causal role in explaining how she came to have the belief. Nevertheless, when we explain how she came to have this belief, it is likely that we would still have to refer back to the number 7 in our explanation and also to the definition of prime numbers. Therefore, it looks like the best explanation of her previous belief about 7 will have to refer to the number itself even if that number is causally inert.

It can then be argued that, in the relevant respects, Anna's belief that 7 is a prime number is similar to our moral beliefs. Following this line of reasoning, non-naturalist moral realists could argue that we also came to have our moral beliefs in a non-causal, non-empirical way where the moral truths, even if they are causally inert, still play an

[49] For an objection to this view of numbers, see Callard (2007). For replies to Callard's objection, see Azzouni (2008) and Park (2019).

explanatory role. They could further argue that this non-causal way of coming to have moral beliefs is based on a priori reasoning. Let us then consider what this non-causal, a priori way could be. According to some non-naturalist moral realists, we come to form moral beliefs through the likes of rational intuition, intellectual intuition, rational insight, intellectual seeming, and so on (Bedke 2009: 188-189; Klenk 2017: 785-786). For the sake of discussion, in this sub-section, I merely consider the non-naturalist moral realists who embrace moral intuitionism, which is the predominant view of how we could come to form moral beliefs in an a priori way.[50]

Some propositions strongly appear to be true. Moreover, it seems that they appear to be true whenever we understand the terms of these propositions well enough. Let us recall the example of Anna and her belief that 7 is a prime number. We can also consider the proposition that 7 is a prime number. This proposition strongly appears to be true when we really understand the terms of that proposition – number 7 and the idea that prime numbers cannot be formed by multiplying two smaller natural number together. Hence, if Anna really understands the terms of the previous proposition, she will also come to have the belief that 7 is a prime number. Nothing else is further needed for Anna to come to have that belief. In this situation, even though numbers are generally considered to be causally inert, the number 7 can still play an explanatory role of how Anna comes to have the previous belief. This is because Anna can rely on her understanding of the causally inert number when she came to have that belief.

[50] Moral intuitionism is usually thought to be the view that "certain cognitions non-inferentially and *prima facie* justify corresponding ethical beliefs" (Bedke 2009: 188). For further discussions of this view, see Audi (2004; 2008), Stratton-Lake (2016), and Wedgwood (2007: 225-247). For a further discussion of why non-naturalist moral realism and moral intuitionism are compatible, see Ridge (2019: Section 3).

We can then turn to the cases of moral propositions and beliefs. Some moral propositions also strongly appear to be true. Let us consider the moral proposition that wanton killing is wrong as an example. It seems that this moral proposition strongly appears to be true when we really understand the term of this previous proposition (Bedke 2008: 254; 2009: 193). We are then strongly inclined to have the moral belief that wanton killing is wrong, and nothing more is required for us to come to have this moral belief. Furthermore, according to the internalism about justification, when the relevant moral proportion is self-evident and we also came to believe it, on the basis of understanding the relevant terms, then nothing more is also required for us to justify that moral belief.[51]

Hence, non-naturalist moral realists could embrace moral intuitionism in order to explain how we come to form moral beliefs in an *a priori* way. On the resulting view, moral truths are considered to be causally inert and yet the moral fact that wanton killing is wrong, for example, can still play an explanatory role of how we come to have the previous moral belief. This is because, according to those non-naturalists, we can rely on our understanding of the causally inert moral truths when we explain how we came to have the previous moral belief (Shafer-Landau 2012: 30).

In response to this objection, I argue that the previous kinds of accounts are not sufficiently good non-causal explanations of our moral beliefs. Our understandings of

[51] In the sub-section §4.4, I will have a further discussion of internalism and externalism about justification. So far in this thesis, I have assumed that internalism is right.

the moral propositions must be, after all, physical in nature (Stratton-Lake 2016: 28; Bedke 2009: 189). This also means that the alleged non-naturalist moral truths would be causally isolated from these physical states of understandings (Ridge 2019: Section 3, para. 5). If the non-naturalist moral realists then wanted to embrace moral intuitionism, they would have to be able to explain how our fully physically caused understandings could accurately represent the non-causal moral truths (Bedke 2009: 190, 196). Non-naturalist moral realists thus face an extra explanatory burden: If they fail to provide a plausible explanation of the previous connection between our physically realized understandings and the non-natural moral facts, then it would still seem like a coincidence that our moral beliefs are true, if they are true.[52]

To sum up, I have first argued that the causal version of Explanatory Account of truth-tracking cannot be used for the purposes of the EDA in 3.2.1. Then, I introduced the liberal version of the Explanatory Account can be used in 3.2.2. In that sub-section, I also suggested that, in accordance with this liberal version, our moral beliefs do not track the moral truths due to the evolutionary origins of these beliefs. Nevertheless, there is also another potential reading of truth-trackingness that could also be used in the EDA. In the next §3.3, I will introduce the modal reading of truth-trackingness and explain how this reading too could be used in the EDA.

3.3 2nd Theory of Truth-Tracking: The Modal Account

[52] It is not necessary that non-naturalist moral realists must accept moral intuitionism. For instance, some non-naturalists instead accept process reliabilism (Shafer-Landau 2003: 267-302) or rely on the method of reflective equilibrium (Parfit 2011: 544; Scanlon 2014: 69-104). For an objection to those non-naturalists who accept process reliabilism, see Lutz (2015: 61-85). For an objection to those non-naturalists who rely on reflective equilibrium, see Hayward (2018).

In this §3.3, I will introduce a different theory of what it is for beliefs to track the truth, that is, the Modal Account of truth-trackingness. According to this account, a given set of beliefs in a certain domain tracks the truths of that domain – including its facts and properties – if there is a certain kind of a counterfactual connection between those beliefs and the relevant truths. §3.3 thus consists of three parts. Firstly, the following sub-section 3.3.1 will introduce the basic crux of the Modal Account. Secondly, in 3.3.2, I will explain what kinds of counterfactual connections could be argued to be required for those beliefs to track the truths on this view in the sub-section. Finally, in the sub-section 3.3.3, I will explain how this Modal Account too can be used by the evolutionary debunkers in the EDA.

3.3.1 The Basic Crux of the Modal Reading of Truth-trackingness

The Modal Account was arguably first intended to replace the causal analyses of knowledge (Kripke 2011: 162). Hence, in order to understand the Modal Account of truth-tracking, we should start from considering a variant of such a view. The defenders of this variant of a causal view of knowledge suggest that, when we discuss how causation is linked to knowledge, we first need to consider what causation actually is. They then come to conclude that causation should be understood in terms of counterfactual relations.

In his famous article "Causation", David Lewis (1973) suggests that we should understand causation in terms of counterfactuals.[53] Since Lewis's ground-breaking

[53] Nevertheless, as David Lewis (1973) himself also suggests, the first philosopher who attempted to define causation in terms of counterfactuals was David Hume (556). For a Humean definition of

work, many other philosophers too have come to rely on counterfactuals to understand and define causation (Bernstein 2019; Ramachandran 1997). The basic crux of the counterfactual theories of causation can be formulated as follows: An event *e* causally depends on another event *c* if and only if, if *c* had occurred then *e* would have occurred; and if *c* had not occurred then *e* would not have occurred (Lewis 1973: 563).

To illustrate this view, we can recall the example of my belief that there is a pen on my desk. Let us also assume that there is a genuine causal connection between the pen on my desk and my belief. On Lewis's view, this amounts to it being true that, if there were a pen on the desk, then I would come to have the previous belief and if there were no pen on the desk, then I would not come to have this belief. Hence, the pen on the desk plays a significant explanatory role with regards to my previous belief by making a real difference to whether the given belief is formed or not. Because of this significant explanatory role, the best explanation of how I came to have the relevant belief must refer to the relevant fact about the pen that obtains in reality and the causal role that the fact plays in the generation of my belief. Since the best explanation of how I came to have the relevant belief must refer to the relevant fact about the pen, the truth-tracking causal condition is satisfied and thus my belief that there is a pen on my desk is tracking the relevant truth. This is how the counterfactual variant of the causal version of the Explanatory Account suggests that the truth-tracking condition can be satisfied in virtue of counterfactual relations.[54]

causation, see Hume (1748/2007: 44-57). For an overview of the views which understand causation in terms of counterfactuals, see Collins et al. (2004).

[54] For other philosophers who attempt to understand causation in terms of the probability theory, see Menzies (2001) and Schaffer (2001).

Let us then consider the Modal Account of truth-tracking. This account too is based on counterfactual relations even though it does not rely on the notion of causation. According to this account, for our moral beliefs to track the relevant moral truths, they must be counterfactually dependent upon the relevant moral truths. And, in addition to the previous counterfactual dependency relation, no further explanatory connections (or causal relations) are required to obtain between our moral beliefs and the relevant moral truths.

Why would the explanatory connections between the moral beliefs and the relevant moral truths not be required for the former to track the latter, according to the Modal Account? This is because, as previously mentioned, if our beliefs are tracking the truth, then our beliefs are formed in a way such that they "adequately align" with the corresponding truths (Bedke 2014: 108; Kahane 2011: 105). The core idea of the Modal Account is that the adequate alignment requires only certain kinds of counterfactual relations between the beliefs and the relevant truths, instead of the explanatory and causal connections. On this view, only the counterfactual relations between the beliefs and the relevant truths can ensure that the processes producing our beliefs can reliably produce true beliefs, or in other words, track the truth (Bedke 2014: 105).

3.3.2 Sensitivity and Safety – Two Potential Kinds of Counterfactual Relations

In this sub-section 3.3.2, I will introduce two kinds of counterfactual relations that could be thought to be required for our moral beliefs to track the moral truth – sensitivity and

safety. The defenders of the Modal Account can hold that either one of these relations is required or they can argue that both relations are required for truth-trackingness. Therefore, there are three potential versions of the Modal Account in the moral context: A moral belief tracks the relevant moral truths (i) if and only if the moral belief is sensitive; or (ii) iff. the moral belief is safe; or (iii) iff. the moral belief is both safe and sensitive. In this sub-section, I will not judge which of these versions of the truth-tracking condition should be accepted by the proponents of the Modal Account. Instead, the remaining part of this sub-section 3.3.2 will explain why the truth-tracking moral beliefs could be thought to require safety and/or sensitivity.[55]

Let us start from considering the sensitivity condition. There is more than one possible formulation of this condition, but Robert Nozick (1981) was one of the first philosophers to introduce it (Jong & Visala 2014: 244n1). He defines the sensitivity condition as follows: "[I]f *p* weren't true and S were to use M to arrive at a belief whether (or not) *p*, S wouldn't believe that *p*" (where *p* is a proposition; S is an agent; M is a method or way of coming to arrive at a belief) (Nozick 1981: 179).

However, as Justin Clarke-Doane (2015) suggests, "the present notion of sensitivity is different from that of Nozick" (88n19). This is because Nozick's definition of sensitivity condition needs to be "relativized to a method of belief formation" (Clarke-Doane 2015: 88n19).[56] Clarke-Doane argues that whether a belief is sensitive or not

[55] The sensitivity and safety conditions were originally considered to be candidates for necessary condition(s) of knowledge in epistemology. See Kelp (2009) and Pritchard (2012a).
[56] Duncan Pritchard (2008) suggests that, if the principle of sensitivity is relativized to the method of belief formation, then the proponents of this principle also face a difficulty to deal with the case like the brain in a vat scenario (BIV) (444-446). For an extensive discussion of this problem, see Becker (2008; 2012).

does not necessarily depend on the method of belief formation. Instead, the sensitivity condition should be essentially defined merely in virtue of counterfactuals. That is, an agent S's belief that *p* is sensitive if and only if, had the relevant truths about *p* been different, S's belief that *p* would have also been correspondingly different (Clarke-Doane 2015: 87-88).

Hence, in this Chapter 3, I will understand the sensitivity condition in the following way: S's belief *p* is sensitive if and only if, if *p* were false, S would not believe that *p* (Joyce 2016b: 129). We can recall the previously mentioned example of my belief that there is a pen on the desk. Assume that there really is a pen on the desk. If the pen were not on the desk, then I presumably would not believe that the pen is on the desk. Therefore, in this example, my belief that there is a pen of the desk is sensitive.

Let us also consider the fake barns case as an example again. When there are no fake barns nearby, Harry will never believe that there is a barn in the field next to him when there isn't one. Hence, in this context, Harry's belief that there is a barn will be sensitive. In contrast, let us consider another situation where there are many fake barns around. In this situation, Harry will believe that there is a barn in the field next to him even when there isn't a real barn around, and as a result, Harry's belief that there is a barn will be insensitive instead.

We can then move on to consider the safety condition. Ernest Sosa (1999) first introduced the safety condition as a potentially necessary condition for knowledge. The proponents of the safety condition believe that a certain kind of a modal connection

between certain truths and our beliefs about them is required for knowledge. They, however, reject the view that the sensitivity condition could really capture this connection.[57] Therefore, they offer the safety condition as an alternative modal condition that could be argued to serve as the relevant modal necessary condition for knowledge.

Sosa (1999) defines the safety condition in the following way: S's belief is safe if and only if, "as a matter of fact, though perhaps not as a matter of necessity, not easily would S believe that *p* without it being the case that *p*" (142). This definition reflects the general spirit behind the safety condition: In cases of knowledge, we do not only get things right as a matter of fact, but we also could not easily get things wrong either (Hirvelä 2019: 1167; Pritchard 2016: 35).

Some epistemologists have also tried to state the safety condition in terms of possible worlds. For example, Richard Joyce (2016b) defines the safety condition in the following way: S's belief is safe if and only if, "in all nearby worlds where S believes that *p*, then *p*" (129). This definition of safety condition also reflects the general idea behind the safety condition. If S's belief is not only true in the actual world but also true in all nearby worlds, then this belief could not have easily been wrong (Pritchard 2016: 27-28). I will also keep this general spirit of the safety condition in mind in the rest of this thesis.

[57] For example, Saul Kripke (2001) uses a modified example of fake barns – the red barn case – to argue that the sensitivity condition is not acceptable. In addition, Keith DeRose (1995) argues that, if we accept the sensitivity condition of knowledge, this also means that we need to accept some problematic "abominable conjunctions" (27-33).

Let us return to the fake barns again. We can first assume that there are no fake barns in the area where Harry is in. In this context, Harry's belief can be thought to be safe. This is because, in all nearby worlds where Harry believes that there is a barn next to him, there will really be a real barn in the field next to him (assuming that in the nearby worlds there are no fake barns in the field). In contrast, let us then consider the situation where there are many fake barns nearby. In many of the close worlds where Harry believes that there is a real barn next to him, there isn't one but rather merely a fake barn. As a result, Harry's belief would be false in these worlds. Therefore, in the fake barn context, Harry's belief that there is a barn next to him fails to be safe even if that belief is true.[58]

3.3.3 Modal Reading of Truth-trackingness and the Evolutionary Debunkers

Let us then consider how the evolutionary debunkers could use the modal notion of truth-trackingness to argue that our moral beliefs are not tracking the moral truth. The basic crux of their argument can be formulated as follows:

The Evolutionary Off-Track Argument (Modal Version)

1. The origin of our moral beliefs can be wholly explained by natural selection (the conclusion of Chapter 2).

[58] For some representative counter-examples to the safety condition, see Juan Comesaña's "Halloween Party" example (2005: 395-404) and Christoph Kelp's "Russell's stopped clock" example (2009: 27-28). For an objection to Comesaña and Kelp, see Bogardus (2014).

2. If the origin of our moral beliefs can be wholly explained by natural selection, then our moral beliefs would be the same, no matter what the moral facts and properties were.

3. Our moral beliefs would be the same, no matter what the moral facts and properties were (1, 2, *Modus Ponens*).

4. If our moral beliefs would be the same no matter what the moral facts and properties were, then our moral beliefs are not sensitive.

5. If the origin of our moral beliefs can be wholly explained by natural selection, then our moral beliefs could have been easily false.

6. Our moral beliefs could have been easily false (1, 5, *Modus Ponens*).

7. If our moral beliefs could have been easily false, then our moral beliefs are not safe.

Conclusion: Our moral beliefs are not tracking the moral truth because our moral beliefs are neither safe nor sensitive (3, 4, *Modus Ponens*) (6, 7, *Modus Ponens*).

This argument is clearly valid. In Chapter 2 and also in sub-section 3.2.2, I have already explained why the premise 1 should be thought to be true. In the rest of this sub-section

3.3.3, I will focus on explaining the premises 2-7 one after another. Let us start from the premise 2. In the previous Chapter 2, I already explained why moral facts and properties are not needed in the evolutionary explanation of the origins of our moral beliefs (as long as either one of the Adaptation Account or Exaptation Account is true). If moral facts and properties are not needed in these evolutionary explanations, then they could also be considered to be irrelevant to how we came to have our moral beliefs through the relevant evolutionary processes (Street 2016: 322).

Let us consider the moral belief that wanton killing is morally wrong as an example again. We can assume now that we can wholly explain the evolutionary origin of this moral belief. This moral belief, of course, is typically thought to be true. Nevertheless, even if this moral belief were false, we still would have come to have the previous belief. This is because moral facts and properties are irrelevant to how we came to have this moral belief in the first place. As long as the origin of this moral belief can be explained in virtue of natural selection, then we would still have evolved to have the same moral beliefs, no matter what the moral facts and properties were, given that those facts and properties play no role in the acquisition of our moral beliefs. For one, it is reasonable to think that the prohibition of wanton killing can increase the survival rate of human beings.[59] This is why we should think that the premise 2 is true.

If the premises 1 and 2 of the argument in question are both true, then the premise 3 will be true as well by *modus ponens*. Let me then explain the premise 4 and why we

[59] I also provided an evolutionary explanation of our moral belief that the wanton killing is wrong in terms of exaptation in the sub-section 2.4.2 above.

should accept it. If the premise 3 is true as I have argued, our moral beliefs would be the same, no matter what the moral facts and properties were. We can return to the moral belief that wanton killing is wrong as an example. By definition, this moral belief is sensitive if and only if, if wanton killing were not wrong, then we would not hold that moral belief. As I just explained, however, the origin of this moral belief can be explained in terms of natural selection. This means that we would still have the same moral belief even if wanton killing were not wrong, and, as a result, this moral belief we have could be claimed to be insensitive. More generally, if our moral beliefs would be the same due to their evolutionary origin no matter what the moral facts were, then these beliefs would be insensitive.[60]

Let me then explain the premise 5 of the argument. According to this premise, if the origin of our moral beliefs can be wholly explained by natural selection, then our moral beliefs could have been easily false. For this premise to be true, it needs to be the case that, if we can provide a genealogical explanation of our moral beliefs and the relevant genealogical factor could have easily been otherwise, then we could have easily come to have false moral beliefs in some nearby worlds. In order to motivate this idea, let's start from non-evolutionary examples. Most people in modern western societies have the following moral beliefs:

- Polygamy is wrong
- Eating dogs is wrong

[60] In the sub-section 3.4 right below, I will introduce and evaluate an objection to this view that our moral beliefs can be thought to be insensitive because of their evolutionary origin.

- Forced marriage is wrong
- Responsible gambling is not wrong
- Responsible drinking is not wrong

Many people in other non-western societies, however, have the polar opposite moral beliefs. For example, it is well-documented that polygamy is not considered to be wrong in many non-western societies and countries even today (Smearman 2009: 385-386). It is also not hard to imagine that any one of us could have been a member of a different society, and thus, as individuals, we could have belonged to a completely different culture.[61] This means that, even if the moral beliefs listed above were true, we could have easily had different moral beliefs due to the cultural variation. Because of the cultural origins of these moral beliefs, we could easily have had many false moral beliefs, and thus many of our moral beliefs are actually unsafe.[62]

We can now consider the evolutionary origin of our moral beliefs and safety. As argued in the previous Chapter 2, we can wholly explain the origins of our moral beliefs in evolutionary terms. Moreover, in that chapter, I also argued that evolutionary debunkers can explain how evolution has influenced our moral beliefs to have certain core contents. We may then ask: Even if we could easily have had many different moral beliefs due to the cultural variation, what about those evolutionarily influenced moral beliefs with

[61] Here, I do not merely consider the thought that we could have been born in a different culture. Moreover, I also include cases where we could have been influenced by many other cultures for different reasons. For example, we could have been adopted by foreign parents, or we could have embraced a different culture because of the impact of the internet.

[62] Here I am relying on the general definition of the safety condition mentioned in the previous sub-section 3.3.2. In the sub-section 3.4 below, I will introduce an alternative definition of the safety condition. According to that alternative safety condition, the listed moral beliefs cannot be claimed to be unsafe, or so I will argue in that sub-section.

core contents that are shared across different cultures? Nevertheless, even though the conventional moral beliefs with core contents are commonly shared by the members of most communities and societies, our moral beliefs could still have been easily different under natural selection.

Let us consider the moral belief that killing a child is wrong as an example again. Regarding infanticide to be wrong is usually thought to be an example of the core contents because the previous moral belief is commonly shared by the members of almost all communities and societies. Yet, it is plausible to think that natural selection could have easily led to different results than it actually did. For example, if our ancestors lived under a different circumstance, they could easily come to have moral belief that is different from the belief that killing a baby is wrong. This thought is supported by the evidence that there reportedly have been at least some cultures which have permitted infanticide (Joyce 2016b: 132). The example often used at this point is Eskimos, who have been known to kill to sometimes kill their newborns in the past, not because of fun, but rather because they lived in an environment where they just would not have been able to feed more people.

Following the same line of reasoning, if our ancestors had also lived under that kind of circumstances without sufficient resources to feed their children, they could have easily come to have moral beliefs according to which killing a baby is not wrong. This means that, even if some of our moral beliefs shared certain core contents universally, those moral beliefs could have been very easily different due to (even small differences in) circumstances. As Richard Joyce (2016b) concludes, as long as there is an uncertainty

over whether our moral beliefs could easily have been different as a result of natural selection, it seems that these beliefs, even if actually true, should be thought to be such that they could have been easily false and thus not safe (132). The premise 5 of the argument can be therefore considered to be true (I am fully aware that there are objections to the premise 5, but I will return to one of them in §3.4).

Given that the premises 1 and 5 of the argument in question are then both true, the premise 6 will also be true by *modus ponens*. Finally, as I already explained above in the previous 3.3.2, moral beliefs are not safe when they could have been easily false, which just is the idea of the premise 7 in the argument above. Hence, the premise 7 of the argument should also be thought to be true. If the premises 6 and 7 of this argument are both true, this also means that the moral beliefs that we commonly have can also be considered as unsafe by *modus ponens*.

If our moral beliefs are insensitive and unsafe, then the conclusion that our moral beliefs are not tracking the moral truths also turns out to be true. Despite this, in the next §3.4, I will focus on one key objection to the previous modal version of the Evolutionary Off-Track Argument. The evolutionary debunkers will need a plausible reply to this objection if they want to use the Modal Account of truth-tracking in the EDA, or so I will argue next.

3.4 The Modal Security Objection and Debunkers' Replies

§3.4 will focus on a recent objection to the Evolutionary Off-Track argument outlined in the previous sections. I will call this objection, which was first outlined by Justin

Clarke-Doane (2012; 2015; 2016; 2017a; 2017b), the "Modal Security Objection". The following sub-section 3.4.1 explains the basic crux of this objection. In 3.4.2, I will then introduce two responses to the objection.

3.4.1 The Basic Crux of the Modal Security Objection

Clarke-Doane suggests that the crucial principle of modal security can be defined as follows: "Information, E, cannot undermine our D-beliefs without giving us some reason to believe that our D-beliefs are not both safe and sensitive" (Clarke-Doane 2015: 97; 2017a: 208). What the evolutionary debunkers attempt to do is to provide an evolutionary explanation of our moral beliefs. By doing so, they believe they can also show that our moral beliefs are neither safe nor sensitive. However, Clarke-Doane thinks that such evolutionary explanations are not a reason to believe that our moral beliefs are unsafe or insensitive (and thus not truth-tracking). In other words, moral realists can still "explain the reliability" of our evolutionarily influenced moral beliefs by arguing that these beliefs are both safe and sensitive (Clarke-Doane 2016: 30-31).[63]

Clarke-Doane grants that the evolutionary debunkers can provide an explanation of our moral beliefs in which no moral truths are mentioned. How could our moral beliefs still be safe and sensitive? This is because, according to almost all ethicists, moral truths are

[63] Clarke-Doane actually uses this principle of modal security to argue that the reliability of our moral beliefs cannot be undermined by their evolutionary origin. However, I believe that his argument can also be used to argue that our moral beliefs do track the moral truth reliably despite their evolutionary origins. This is because Clarke-Doane also argues that the evolutionary debunkers fail to provide a reason why our moral beliefs would not be both safe and sensitive (Clarke-Doane 2015: 97; 2017a: 208). If our moral beliefs are both safe and sensitive, then our moral beliefs can be considered to be tracking the moral truth. Hence, Clarke-Doane's argument can also be considered to be an objection to the Evolutionary Off-Track Argument.

necessary truths (Clarke-Doane 2012: 320; 2016: 32; Silk 2013: 196-197).[64] If we want to deny this claim, the problem is that then we are also required to reject moral supervenience (Väyrynen 2018: 179-182).[65]

To illustrate this, let's again consider the moral belief that wanton killing is wrong. Let us also assume that act-utilitarianism is true in the actual world (Väyrynen 2018: 181; Rosen forthcoming). In the actual world, the moral belief that wanton killing is wrong would thus be true because the act of wanton killing never maximizes the total amount of happiness. Let us then consider the possibility that moral truths would not be necessary truths. In this case, there could be a metaphysically possible world W where the non-moral features of this world W are exactly the same as in the actual world, and yet act-utilitarianism is false. This means that, in this possible world W, the moral belief that wanton killing is wrong could be false.

Let us then consider two persons – Tim in the actual world and Tom in the possible world W – as an example. In the actual world, Tim killed many innocent people, and so he has done many morally wrong actions. It is intuitive to think that if what Tim does is morally wrong, then anyone who would do exactly the same thing would also be acting wrongly (Väyrynen 2018: 179). Nevertheless, Tom would not be doing anything

[64] Some realists argue that moral truths are not just metaphysically necessary truths but also conceptually necessary truths (Cuneo & Shafer-Landau 2014; Clarke-Doane 2016: 36n21). I will make a separate objection to those realists who think that moral truths are conceptually necessary truth in Chapter 5.

[65] Many moral truths are apparently contingent. Mark Schroeder (2014) suggests that we need to make a distinction between pure and "bastard" moral claims (130-131). A bastard moral claim is a conjunction of a pure moral claim and a non-moral claim. If a bastard moral claim is true, it is also contingently true because the non-moral part of the bastard moral claim must be contingently true. Hence, the reason why we think that many moral truths are apparently contingent is because these moral truths are bastard moral truths.

wrong in the world W, even if he acted exactly in the same way as Tim does in the actual world. This is because we are assuming that, because moral truths are not necessary, the moral truths of W are different from in the actual world even if the two worlds are otherwise alike. The price of denying that moral truths are necessary truths is that you have to deny moral supervenience as well.

In contrast, if we accept that moral truths are necessary truths, then we do not need to deny moral supervenience. If it is true that wanton killing is wrong necessarily, then wanton killing would be wrong in all possible worlds. Whatever possible world we take, if you killed as many innocent people in that world as Tim did in the actual world, you would also be acting wrongly. This is why moral truths are thought to be necessary truths.[66]

We can then return to Clarke-Doane's argument to the conclusion that our moral beliefs must be sensitive. If moral truths are necessary truths, then it could be argued that many of our moral beliefs are necessarily true. We can then return to the idea of sensitivity: One's moral belief that *p* is sensitive if and only if, if it were not-*p*, one would not believe that *p*. If it is necessary that *p*, then it is impossible that not-*p*. Hence, the antecedent of the previous counterfactual condition, i.e., "if it were not-*p*", must be necessarily false. It is generally accepted in logic that any subjunctive conditional with

[66] Some philosophers reject the view that moral truths are necessary truths in the context of the problem of moral supervenience. See Fine (2002) and Rosen (forthcoming). Nevertheless, even if they succeeded, their objections to the view that moral truths are necessary truths are unavailable to the evolutionary debunkers. This is because the EDA is considered to be an epistemological objection instead of a metaphysical objection (Clarke-Doane 2016: 36n23). As I mentioned in Chapter 1, my thesis has been mainly focusing on the epistemological versions of the evolutionary debunking arguments. It is beyond the scope of this thesis to argue against the metaphysical view that moral truths are necessary truths.

a false antecedent must be true. Hence, according to this principle, the right-hand side condition of the definition of sensitivity (i.e., if it were not-*p*, one would not believe that *p*) will always be satisfied in every case of true moral beliefs. This is because this condition consists of a counterfactual conditional of which the antecedent must be false, given that moral truths could not have been otherwise in any other possible worlds. Therefore, all moral beliefs about moral truths are trivially and vacuously sensitive (Jonas 2017: 2735; Clarke-Doane 2017a: 206).

Clarke-Doane also considers a possible objection from the evolutionary debunkers. He thinks that evolutionary debunkers could argue that many of our moral beliefs are actually false (Clarke-Doane 2012: 320; 2017a: 206-207). As I have previously mentioned, members of different cultures often have completely opposite moral beliefs. It is unlikely that all the beliefs of our culture are true, whereas the beliefs of the other cultures are all false. Let us then consider a moral belief that p, which happens to be false, as an example. In this situation, the previously mentioned antecedent, "if it were not-*p*", will not be necessarily false (but rather necessarily true). Thus, the counterfactual condition that "if it were not-p, one would not believe that *p*" will not be trivially true either. Hence, it is possible that this moral belief p could be insensitive. As a result, debunkers could be thought to argue that at least some of our moral beliefs could be insensitive because they are actually false.

In reply to this possible objection from the evolutionary debunkers, Clarke-Doane argues that some of our "explanatorily basic moral beliefs" should be considered to be necessarily true (Clarke-Doane 2015: 88-90; 2016: 26; 2017a: 207). These basic moral

beliefs are most likely to be some of our core, everyday moral beliefs, including the moral belief that wanton killing is morally wrong. It is impossible for us to imagine that this moral belief could be false, and likewise it is also impossible for us to imagine that the relevant moral truth could be different (Clarke-Doane 2012: 321).[67] As a result, Clarke-Doane argues that evolutionary debunkers fail to show that all of our basic moral beliefs are false (and thus at least some of these beliefs must be trivially sensitive).[68]

Finally, we can then consider safety. Let's also assume that Clarke-Doane is right that moral truths are necessary truths and that some of our basic moral beliefs are necessarily true. He also grants that we can provide an evolutionary explanation of how we came to have our moral beliefs. Because of these evolutionary origins of our moral beliefs, however, it could be argued that we could not have ended up having different moral beliefs easily (Clarke-Doane 2015: 93; 2016: 28; 2017a: 207). At least, we could not easily have had different core moral beliefs.

Let's consider all nearby possible worlds where the non-moral facts are very similar to our actual world. The evolutionary influence on our moral beliefs in these nearby worlds would be very similar to the influence on our moral beliefs in the actual world.

[67] Some philosophers thus suggest that Clarke-Doane is arguing that these core moral beliefs are "evolutionarily inevitable" (Tersman 2016: 42), "determined" and "hard-wired" (Jonas 2017: 2735-2736).

[68] However, if the evolutionary debunkers really wanted to argue that our basic moral beliefs (i.e., our core and everyday moral beliefs) are actually false, this would also mean that they would be required to show that the moral error theory is generally true. After all, only the moral error theorists would hold the view that all our core and everyday moral beliefs are actually false. For a representative proponent of the moral error theory, see Mackie (1977). Nevertheless, if the moral error theory is really true, this also implies that the EDA is redundant. According to the moral error theory, all moral beliefs (including those basic moral beliefs) are "systematically and uniformly false" (Fisher 2011: 46). If this were the case, there would not be a genuine need to use the EDA to show that the epistemic status of our moral beliefs is undermined due to their evolutionary origin. As a result, it would be unreasonable for the evolutionary debunkers to hold a meta-ethical view which would make the EDA redundant.

Hence, the evolutionary origins our moral beliefs would be almost the same in the actual and the nearby worlds. Therefore, we would come to have the same moral beliefs in these possible worlds and in the actual worlds for the very same evolutionary reasons.

Moreover, as we saw above, there are reasons to believe that moral truths are necessary truths, and as a result the moral truths are also exactly the same in the actual world and the previous nearby worlds. This means that the moral beliefs that are true in the actual world are also true in these nearby worlds. Let us then recall the definition of safety in terms of possible worlds: Our moral belief that *p* is safe, if and only if, in all nearby worlds where we have this moral belief that *p*, *p* is true. As our basic moral beliefs would be true in the nearby possible worlds and they are exactly the same as our moral beliefs in the actual world, they are also safe, or so Clarke-Donae argues.[69]

Let us consider the core moral belief that wanton killing is wrong again as an example. We all think that this moral belief is true in the actual world. And, arguably, we would evolve to have this same moral belief in all the nearby possible worlds, where the evolutionarily influence on the moral belief is similar to the influence evolution exerts in the actual world. Therefore, in all the relevant nearby worlds, we would still come to have the same moral belief that wanton killing is wrong, and this belief would still be true in these worlds, given that it is a necessary truth. As a result, this moral belief is safe. Following the same line of reasoning, all basic moral beliefs that we would still have in all the nearby possible world due to evolution would also be true beliefs in those

[69] Note that, according to this definition of safety, it makes no difference whether we could easily have had different moral beliefs. This also means that Clarke-Doane must be relying on the notion of safety which is very similar to the general idea of safety that I mentioned in the sub-section 3.3.2 above.

worlds (Clarke-Doane 2017b: 35). As a result, these basic moral beliefs are safe.

To sum up, if Clarke-Doane is right, our moral beliefs are tracking the moral truth. This is because our moral beliefs are safe and sensitive even if these beliefs have an evolutionary origin.

3.4.2 Replies

This sub-section 3.4.2 will outline two ways in which evolutionary debunkers can respond to the Modal Security objection. I call the first response 'the explanatory connection objection' and the second one 'the anti-luck requirement objection'. In this 3.4.2, I will argue that these two objections can help evolutionary debunkers to avoid the Modal Security Challenge.

Both of my objections originate from the same worry. The concern is that, even if some of our moral beliefs were safe and sensitive in the way described by Clarke-Doane, these beliefs would still not track the moral truth. To see this, let us return to the example of the belief that 7 is a prime number. Now, it is of course a necessary truth that 7 is a prime number. Hence, when I have this belief that 7 is a prime, my belief will also be necessarily true. Moreover, my belief that 7 is a prime number is also trivially sensitive and safe for the following reasons.

Let us consider sensitivity first. As mentioned in 3.3.2, my belief that 7 is a prime number is sensitive, if and only if, if 7 were not a prime, I would not have this belief. However, given that 7 is a prime number in all possible worlds, the antecedent of the

previous counterfactual conditional (i.e., "if 7 were not a prime") must be necessarily false. As I already explained in 3.4.1, this entails that the right-hand side condition of the definition of sensitivity (i.e., if 7 were not a prime, one would not have this belief) will always be satisfied. Hence, my belief that 7 is a prime is trivially sensitive.

Let us then consider safety. According to the definition of safety in terms of possible worlds (as described in 3.3.2), my belief that 7 is a prime number is safe if and only if, in all nearby worlds where I have this belief, 7 really is a prime number. Given that it is a necessary truth that 7 is a prime number, 7 is a prime number in all possible worlds. Therefore, my belief that 7 is a prime is true in the actual world and all nearby worlds where I have this belief that 7 is a prime number. Hence, this belief that 7 is a prime too is trivially safe.[70]

Nevertheless, we can now consider another person, Alex, who came to have this belief that 7 is a prime number merely on the basis of rolling a dice and not on the basis of understanding the relevant concepts or anything like that (Joyce 2016b: 133). Nevertheless, his belief must be necessarily true as 7 is a prime number in all possible worlds. Hence, even in this case, Alex's belief would still be trivially safe and sensitive for exactly the same reasons as my corresponding belief in the previous case. However, Alex was very lucky when he acquired the true belief that 7 is a prime number. If he

[70] As mentioned in 3.3.2, we can understand safety also in a different way: S's belief *p* is safe if and only if S would not easily believe that *p* without it being the case that *p*. However, in this example, my belief that 7 is a prime number is still trivially safe according to that alternative definition of safety too. This is because it is impossible that 7 is not a prime number. Hence, it is also impossible for me to have the belief that 7 is a prime number without it being the case that 7 is not a prime number. Hence, my previous belief could not have been false, and thus it must be safe even according to the alternative definition of safety.

came to have this belief entirely on the basis of the roll of the dice, then the result of the roll of the dice caused him to have this belief. This means that he came to have this true belief merely on the basis of a completely random process. In many nearby worlds, Alex gets a completely different result from the roll of the dice, and thereby he ends up acquiring a false belief in those worlds.[71]

This example nicely illustrates the worry that the safety and sensitivity conditions do not "achieve [their ends] of eliminating epistemic luck" (Collin 2018: 487). Intuitively, if a belief really tracks the relevant truth, this belief should not be true merely by luck. However, in the previous example of Alex, although his belief is safe and sensitive, it is still true by mere luck. As a result, it seems that safety and sensitivity are not sufficient for beliefs to track truth, and so epistemologists have discussed additional requirements.[72] According to different epistemologists, there are two possible options of what the additional requirement could be. The first option is to add a separate explanatory condition (just like what I introduced in the previous 3.2.2). Another option to eliminate the previous kind of epistemic luck is to revise the safety condition. In the rest of this 3.4.2, I will discuss these two options in turn.

(i) The Explanatory Connection Objection

As I explained in §3.1, the general idea behind the notion of truth-trackingness is that

[71] Jaakko Hirvelä (2019: 1168) and Duncan Pritchard (2012b: 256) introduce a similar example: A person uses a malfunctioning calculator that generates random answers, and she accidentally gets the right result of 12 × 13 = 156 by using that calculator (Hirvelä 2019: 1168). Given that [12 × 13 = 156] is a necessary truth, her belief that [12 × 13 = 156] is necessarily true and seems to be trivially safe (Hirvelä 2019: 1168). Besides, James Henry Collin (2018) also introduces a similar "luck 8-ball" example (487).

[72] For example, in Duncan Pritchard's early writing (2005; 2009: 34), he attempted to deal with this problem by trying to restrict the safety condition to contingent truths only. However, he abandons this approach in his later work (Pritchard 2016: 36).

truth-tracking beliefs should be formed in a way that ensures that these beliefs adequately align with the corresponding truths. If these beliefs are adequately aligned with the corresponding truths, then it should also not be an epistemic accident that these beliefs are true. Even though safety and sensitivity can provide certain kinds of alignment between beliefs and the relevant truths, the previous example of Alex shows that those kinds of connections are not sufficient to exclude all epistemic accidents (Setiya 2012: 91). Thus, the previous two modal connections cannot constitute the required kind of adequate alignment between the truth-tracking beliefs and the corresponding truths, contrary to what Clarke-Doane has argued. One possible solution to the previous problem has been that, in addition to the safety and sensitivity condition, truth-trackingness also requires that there is an explanatory connection between the beliefs and the relevant truth (Joyce 2016b: 132-133; Woods 2018: 52).

Let us consider Alex again. Alex's belief can wholly be explained by the die roll. Hence, even if 7 really is a prime number, this fact does not help to explain how Alex came to have his belief (Joyce 2016b: 133). This is because Alex did not acquire the previous belief by relying on the fact that 7 is a prime. In other words, Alex came to have this belief through a belief-formation process that isn't connected in any way to the relevant truth (Joyce 2016b: 133). Thus, there is no explanatory connection between Alex's belief and the fact that 7 is a prime number. This also explains why Alex's belief is not tracking the truth. And so, it could be argued that, according to the requirement of the explanatory connection, this prevents the epistemic luck in this case of necessary truth.

Hence, according to this Explanatory Connection Objection, our moral beliefs track the

moral truth only if there is an explanatory connection between those beliefs and the relevant truths. Furthermore, if the evolutionary debunkers accept the liberal version of the Explanatory Account – which is introduced in the previous 3.2.2 – in the first place, then the Modal Security objection could not pose a real challenge to them. This is because, in that sub-section, I already explained why the requirement of the explanatory connection between our moral beliefs and the moral truths is unlikely to be satisfied due to the evolutionary origin of those moral beliefs. As a result, even if our moral beliefs were safe and sensitive, they would not really track the relevant moral truths, as opposed to what Clarke-Doane has argued.

(ii) The Anti-Luck Requirement Objection

As previously mentioned, I argued that merely modal connections such as safety and sensitivity are not sufficient to ensure that our beliefs track the relevant truths in a way that would eliminate epistemic accidents. In this situation, some epistemologists think that this is because the original conditions of safety and sensitivity are not considered to be the best way to understand the relevant modal condition (Pritchard 2012b: 249-257). They therefore try to introduce an additional modal anti-luck requirement. These philosophers are often called 'anti-luck epistemologists' (Pritchard 2012a: 173; 2016: 18), 'virtue epistemologists' or 'virtue reliabilists' (Pritchard 2016: 44).[73]

[73] As Duncan Pritchard (2016) explains, virtue reliabilism is a form of virtue epistemology (44). The essential idea of the virtue ethics is roughly that the morally virtuous agent "who has the right mix of virtuous moral traits … should be admired and emulated" (Pritchard 2016: 44). Following the same line of reasoning, virtue epistemologists argue that epistemology should also concern some cognitive traits that "one should possess in order to be a 'good' epistemic subject" (Pritchard 2016: 44). On their view, essentially, for a good and virtuous epistemic subject to possess certain knowledge, her cognitive faculties should be somehow involved (and the epistemic luck should thus be eliminated) when she comes to have the relevant knowledge.

The essential aim of the anti-luck epistemology is to accommodate the anti-luck intuition: Knowledge is thought to be excluding luck in the sense that it is not a matter of luck for one's belief to be true (Pritchard 2012a: 173). As argued above, the safety and sensitivity conditions are unable to do this. Hence, in order to accommodate the anti-luck intuition, some anti-luck epistemologists instead attempt to reformulate the safety condition.[74] They introduce the revised safety condition that makes safety relative to a basis, an ability or a method (Greco 2012: 195; Pritchard 2012b: 263) (For the sake of brevity, I will just consider the revised safety condition that makes safety relative to a basis in this chapter). The revised safety condition can thus be formulated in the following way: S's belief that *p* is safe*, if and only,* in most nearby worlds where S believes that *p* on the same basis B that S believes that *p* in the actual world, *p* is true (Greco 2012: 196; Hirvelä 2019: 1168).

Two important clarifications are needed: Firstly, according to Pritchard (2012a), the notion of basis can be roughly defined as "what gave rise to [a] belief" (178).[75] For instance, if I see a pen on the desk and believe that there is a pen on the desk, then my vision is my basis for me to form the previous belief. Secondly, the revised safety condition specifically concerns the nearby worlds where "at the very least the agent forms her belief in the same way as she does in the actual world" (Pritchard 2005: 152). If, in a possible world, the agent was unable to form her belief on the same basis as she

[74] Basically, most anti-luck epistemologists abandon the sensitivity condition and merely attempt to revise the safety condition. As Duncan Pritchard (2016) explains, safety can effectively help us to exclude at least some lucky knowledge (but not all) (31). See also Pritchard (2012a: 177). In contrast, sensitivity is unlikely to help us to exclude any types of lucky knowledge. This is because, as suggested by John Greco (2012), "whether our informants are sensitive to far-off counterfactual situations … should not matter whether they are reliable or dependable in situations vastly different from the contexts in which their information is needed" (199).

[75] For an extended discussion of how bases can be individuated, see Hirvelä (2019: 1172).

does in the actual world, then this possible world should not be considered to be the nearby world that is relevant to safety.[76]

To demonstrate how the revised safety condition works, we can recall the example of Alex's belief that 7 is a prime number. Alex came to have this belief on the basis of rolling a dice. Although his belief is necessarily true, there are also many nearby possible worlds where he forms a false belief about a certain number being a prime on the same basis of rolling a dice (Roland & Cogburn 2011: 557). This is because rolling dice is a completely random process. In other words, Alex could have easily had a different belief that would have been false on the same basis (Hirvelä 2019: 1172). Therefore, according to the revised safety condition, Alex's belief that 7 is unsafe even though his belief is necessarily true. Hence, the essential idea behind the basis-relative safety condition is that we are not only interested in whether a given belief could not have been easily false, but we are also interested in whether an agent could have easily come to have a false belief on the same basis that she formed her belief in the actual world (Hirvelä 2019: 1171).[77]

Let us then consider how the revised, anti-luck safety condition can help the evolutionary debunkers to deal with the problem of the moral beliefs that are about the

[76] Duncan Pritchard (2012a) also invites us to understand safety in terms of avoiding 'risk of error'. As he suggests, "in the very closest nearby possible worlds we are extremely intolerant when it comes to such epistemic risk, and so would not want to be forming any false beliefs on the target basis" (Pritchard 2012a: 178-179). In other words, if we could have easily had different moral beliefs on a given basis in the nearby possible worlds, then there could be a high risk of forming false beliefs on the same basis (Pritchard 2016: 33).

[77] As mentioned in 3.3.2, the general idea behind the safety condition is that we do not get things right only as a matter of fact, but rather we also could not have got things easily wrong either. This revised safety condition thus reflects that general idea behind the safety condition with a significant emphasis on the belief-formation bases.

necessary moral truths. To recall, according to Clarke-Doane, our moral beliefs are necessarily true, and, as a result, it is impossible to imagine that these moral beliefs could be easily different. Nevertheless, as I will argue below, we could easily have arrived at different moral beliefs on the same evolutionarily influenced basis on which we formed our moral beliefs. Hence, even if our moral beliefs were necessarily true, these beliefs could still be unsafe according to the basis-relative safety condition.

To illustrate this view, let us start from considering marine bacteria as a biological, non-moral example. Elliott Sober and David Sloan Wilson (1998) suggest that there are some kinds of marine bacteria that need to avoid oxygen (304-307). If these bacteria fail to avoid oxygen, they will eventually die (Rolfe et al. 1978). Hence, they need to have evolved to have a trait that can help them to avoid the lack of oxygen when they live under the water. Nevertheless, different bacteria react to the very same adaptive problem with different adaptive solutions even under similar circumstances. For example, some kinds of marine bacteria have evolved to have a direct oxygen detector (Sober & Wilson 1998: 304). On the other hand, some other kinds of bacteria have evolved to have magnetosomes, which is a kind of device that can "orient their swimming to the earth's magnetic field" (Sober & Wilson 1998: 304). The concentration of oxygen in deep water is much lower than the concentration in surface water. Hence, oxygen detectors and magnetosomes are two different kinds of marine bacteria's trait to avoid oxygen (Sober & Wilson 1998: 305).

This biological example of marine bacteria thus illustrates that, in the case of evolution, there are no fixed ways for the species to deal with an adaptive problem. When the

organisms of the same kind faced the same adaptive problem, they could have adopted different adaptive solutions under natural selection. This can be argued to be true also in the case of moral beliefs. It has always been the case that different human populations have faced many similar adaptive problems. Nevertheless, we could have arrived at very different moral beliefs on the same belief-formation basis even when facing the same adaptive problems under similar circumstances.

To see this, we can consider our need for avoiding violence as an example. Avoiding violence is often considered to be an essential need for human survival and flourishing. Different societies, however, end up with very different ways of satisfying this need (Gaus 2010: 42). Some societies arrive at the belief that only strict laws supported by strong punishments can prevent violence. For example, members of these societies come to have the moral belief that capital punishment is not wrong. In contrast, many other societies believe that they should regulate violence through education and religion. Hence, members of these societies have the very different moral belief that capital punishment is wrong. In this example, even if members of different societies could have different moral beliefs about capital punishment, they have arrived at different beliefs on the same evolutionarily influenced basis of satisfying the need for avoiding violence.

Many other moral beliefs that we have in the actual world are similar to our moral beliefs about avoiding violence (for example, beliefs about the way that we collaborate with others, beliefs about our relationship with nature, and so on). In the actual world, we come to have certain moral beliefs on the evolutionarily influenced basis. In some nearby possible worlds, however, we could come to have very different moral beliefs

on the same basis (Beatty 1993: 58). Nevertheless, in both actual world and those nearby possible worlds, we arrive at different moral beliefs on the same basis in order to deal with the very same adaptive problem, just like the previous example of avoiding violence. As a result, we could easily have had many different moral beliefs (and thus false beliefs) on the same belief-formation basis. Most of our moral beliefs are therefore not tracking the moral truth because the revised safety condition of these beliefs is not satisfied.

Presumably, there are very few core moral beliefs that could not have been different. That is to say, for a few moral beliefs, we would have the same moral beliefs in both actual worlds and the nearby possible worlds. Evolutionary debunkers could happily concede that perhaps there are only a limited number of core moral beliefs that are safe in this way (for example, the moral belief that wanton killing is wrong). This is shown by the fact that different societies have ended up with very different kinds of moral codes and principles and only share very few common moral beliefs. With the exception of these rare core beliefs, evolutionary debunkers could still argue that most of our moral beliefs are actually unsafe (and thus not tracking the moral truth).[78] This means that it would be a pyrrhic victory for Clarke-Doane to hold that only very few moral beliefs are considered to be safe.[79]

3.5 Conclusion

In this chapter, I have introduced and evaluated two different views of how we could

[78] This would lead to "local" evolutionary debunking arguments (Rowland 2019: Sinclair 2018: 99).
[79] Some non-naturalist moral realists introduce a similar idea of "moral fixed points" (Cuneo & Shafer-Landau 2014). I will make an extended objection to the moral fixed points theory in Chapter 5.

understand of what kind of alignment between beliefs and facts truth-trackingness consists. I have argued that the liberal version of the explanatory reading of truth-trackingness and the modal reading of truth-trackingness are both plausible accounts that can be used by the evolutionary debunkers to argue that our moral beliefs do not track the moral truth. Finally, in the end of this chapter, I have focused on rejecting Clarke-Doane's Modal Security objection to the evolutionary off-track argument.

As a result, there are two different ways that evolutionary debunkers could argue that our moral beliefs do not track the moral truth. If the evolutionary debunkers accept the liberal version of the explanatory reading of truth-trackingness, then they could argue that there is a lack of explanatory connection between our moral beliefs and the relevant truth due to the evolutionary origin of these beliefs. Alternatively, if the debunkers adopt the modal reading of truth-trackingness, then they could rather argue that almost all of moral beliefs are unsafe due to their evolutionary origin, according to the revised safety condition. Either way, the debunkers can also conclude that most, if not all, of our moral beliefs are not tracking the moral truth due to the evolutionary origin of these beliefs. In the next chapter, I will explain how the fact that moral beliefs are not tracking the moral truth can then be used to undermine the epistemic status of our moral beliefs.

Chapter 4

Undermining the Epistemic Status of Moral Beliefs

4.1 Introduction

As already mentioned in the introductory chapter, most proponents of the EDA, if not all, want to argue that the epistemic status of our moral beliefs is undermined due to the evolutionary influence on these beliefs. Yet, we have not yet reached that conclusion here. Rather, so far I have only argued for two intermediate conclusions: (1) Evolution shaped our moral beliefs (through the processes of adaptation and exaptation) and (2) most, if not all, of our moral beliefs are not tracking the moral truth reliably due to the evolutionary origins of these beliefs. However, even if our moral beliefs were not tracking the moral truth reliably, this would not necessarily mean in itself that the epistemic status of these beliefs would thus be undermined. A further explanation of how we could reach that conclusion is needed at this point from the debunkers.

For this reason, in this Chapter 4, I will explain both why and how the epistemic status of our moral beliefs is undermined by the fact that our moral beliefs are not tracking the moral truth due to their evolutionary origin. This chapter consists of three parts. Firstly, §4.2 will introduce three different kinds of defeaters – rebutting, undercutting and higher-order defeaters – that are all often discussed in epistemology.

Secondly, in §4.3, I will first explain why the fact that our moral beliefs are not tracking the moral truth due to their origin fails to constitute a rebutting defeater to our moral beliefs. I will then argue that the dominant account of the undercutting defeaters, which is from John Pollock (1974, 1987), is problematic. To this end, I will introduce Scott Sturgeon's objection (2014) to Pollock's account of undercutting defeaters. After Sturgeon's objection, I will argue that, in order for there to be an undercutting defeater, we need to have a belief that our moral beliefs are not tracking the moral truth due their origin and also a higher-order belief about the evolutionary explanation of the origin of our moral beliefs. The conjunction of those beliefs then jointly provides an undercutting defeater for our moral beliefs. I will then argue that evolutionary debunkers can also translate this EDA-based undercutting defeater into a higher-order defeater for our moral beliefs. As a result, §4.3 concludes that the epistemic status of our moral beliefs is undermined because of these two kinds of defeaters.[80]

In §4.4, I will finally consider a possible objection from the moral realists. They may argue that the epistemic status of our moral beliefs cannot be undermined by any defeaters, given that the externalism concerning justification is true. In reply to this objection, I will argue that the previous fact that our moral beliefs are not tracking the moral truth due to their origins and the evolutionary explanation of the origin of our moral beliefs can jointly show that our moral beliefs are produced by a non-reliable belief-formation mechanism. According to externalists, this means that our moral beliefs have never been justified in the first place. As a result, the epistemic status of

[80] Not all versions of EDA argue that the epistemic status of our moral beliefs is undermined because of defeaters. For example, see FitzPatrick (2015), Klenk (2017), and Leibowitz and Sinclair (2017:213-224).

our moral beliefs would still be undermined even if the externalism were true, or so I will further argue in §4.4.

4.2 Three Kinds of Defeaters

In this sub-section, I will focus on the concept of defeaters in epistemology. What actually is a defeater? There are many considerations that justify our beliefs, but evidence is usually considered to be the most important kind of a justifying consideration. For instance, I see a ball in front of me, and this ball looks green. Based on the evidence of the ball looking green, I am justified in believing that this ball really is green. Nevertheless, it is also very common for us to lose the justification we have for our beliefs, and one common way in which this can happen is through epistemic defeat. Generally, we can come to have new information that can defeat the justification that we had for our beliefs on the basis of the original evidence (Grundmann 2011: 158; McPherson 2020: 10; Lasonen-Aarnio 2014: 314).[81]

How exactly does the new information undermine the justification that we had for our beliefs? There are three different kinds of defeaters: rebutting, undercutting and higher-order defeaters.[82] In accordance with these three different kinds of defeaters, there are also three different ways in which we can lose our initial justification for a belief through defeat (Christensen 2010: 193-199; Easwaran 2015: 146; Janvid 2017: 703; McPherson 2020: 10-11; Lasonen-Aarnio 2014). In the rest of this §4.2, I will introduce

[81] Not all defeaters aim to defeat the justification of beliefs. For example, there are some defeaters that aim to defeat one's knowledge (Pritchard 2018: 3069).
[82] Some epistemologists instead use the term "overriding defeaters" for rebutting defeaters (Pritchard 2018: 3069-3070) and the term "undermining defeaters" for the undercutting defeaters (Janvid 2018: 703n4).

these three kinds of defeaters respectively.

4.2.1 Rebutting Defeaters

Let us start from considering the rebutting defeaters first. According to John Pollock (1987), we may come to have new pieces of evidence that are *prima facie* reasons against our previous beliefs (484). These reasons, which can be considered to be rebutting defeaters, outweigh the evidence that we had for our beliefs. As a result, we lose the initial justifications that we had for our original beliefs. This is why the so-called rebutting defeaters defeat the epistemic justifications that we had for our beliefs by outweighing that justification.

Let's consider a proposition P as an example. Assume that there is evidence E_1 that supports P, and I thus have a reason R_1 to believe P due to E_1. Let's imagine that I then come to have new evidence E_2, and so I also come to have a prima facie reason R_2 to believe ~P due to E_2 (Easwaran 2015: 146). In this case, R_2 is a rebutting defeater. When there is this rebutting defeater R_2, I lose the initial justification that I had for my belief that P. This is because R_2 outweighs the evidence E_1 that I had for my belief that P, and thus I should no longer hold that belief (Klein 2014: 2719; Plantinga 2000: 359). After all, it is inconsistent for me to believe that P and ~P at the same time. If I firmly believe that ~P due to R_2, then I should consider giving up my original belief that P in order to avoid having inconsistent beliefs (Plantinga 2000: 359).[83]

[83] Rebutting defeaters can also be experiences, beliefs and proposition, as long as we can count them as additional evidence against the justification that we had for our beliefs (Grundmann 2011: 158).

We can imagine, as an example, that Chris has been living on a remote island since her birth. Let us also imagine that all dogs are black on the island where she lives. Thus, Chris has always seen black dogs. Based on this evidence, she comes to have the belief that all dogs are black, and arguably her belief is justified on the basis of her having seen only black dogs so far. Let us also assume that her cognitive and perceptual faculties have always been functioning well. One day, however, a foreign sailor brings a white dog to this remote island, and Chris also sees this dog. Based on the new evidence, she comes to have a *prima facie* reason to believe that not all dogs are black. In this case, this *prima facie* reason is a rebutting defeater, as this reason outweighs the evidence that she had for her belief that all dogs are black.

When Chris becomes aware of this rebutting defeater, she is in an epistemic position in which her confidence in her original belief should be reduced. This is because the rebutting defeater is inconsistent with her previous belief. Chris cannot rationally continue to hold her original belief that all dogs are black together with the rebutting defeater, and therefore she loses her initial justification for her original belief and, as a result, she should give up that belief.

4.2.2 Undercutting Defeaters

Let us then turn to the undercutting defeaters. In this 4.2.2, I will explain John Pollock's view of the undercutting defeaters (1987: 484). As I will then explain in §4.3, his view of undercutting defeater is not without its problems (Casullo 2018; Sturgeon 2014). Nevertheless, in this sub-section 4.2.2, I will set aside these problems (I will then discuss one of these problems in 4.3.2 below).

As already explained, we may come to have new information that can defeat the justification that we had for our beliefs. However, when the new information functions as an undercutting defeater, it defeats the justification that we had for our original beliefs in a different way than in the previous case of rebutting defeaters. Let us consider my belief that P as an example again. Let us also assume that there is evidence E_1 that supports P, and I thus have a reason R_1 to believe P due to E_1. Let us then imagine that I come to have a piece of new evidence E_3. E_3, however, is not evidence for ~P. Instead, E_3 is evidence against the idea that my original evidence E_1 in fact supports P. When this is the case, E_3 can be considered to be an undercutting defeater.

As an undercutting defeater, R_3 undercuts the evidence E_1 that I had for my belief that P (Klein 2014: 2719; Plantinga 2000: 359). In other words, E_3 makes it the case that evidence E_1 no longer evidentially supports the belief that P (Lutz 2018: 1109). My confidence in my belief that P should therefore be significantly reduced (Janvid 2017: 703). This is because, based on E_3, I come to have a *prima facie* reason to believe that the initial R_1 (that is, the ground of my justification for my belief that P) is not a good reason to believe P (Easwaran 2015: 146). As a result, I may then lose the justification that I had for my belief that P.

Let us consider the following example. Lucas sees a table in front of him, and that table looks red to him (Christensen 2010: 194; Lutz 2018: 1109). Thus, he comes to have the belief that the table is red on the basis of its visual appearance. Based on that evidence, Lucas is thus justified in believing that the table is red (Christensen 2010: 194). Let's

imagine then that Lucas comes to learn that there is a red light shining on the table (Lutz 2018: 1109). This new evidence, however, cannot provide a rebutting defeater. After all, the table could really be red even if it were illuminated by a red light. In other words, the new information that the table is illuminated by a red light has nothing to do with the proposition that the table is not red. Hence, even if Lucas's original belief were defeated by this new piece of evidence, it is not because his original belief that the table is red is inconsistent with the new evidence.

So how can the justification of Lucas's original belief be defeated in this case? As an undercutting defeater, the new evidence makes it the case that the original evidence that the table looks red no longer evidentially supports his original belief that the table is red. If Lucas becomes aware of the undercutting defeater (i.e., the new evidence), then that original evidence no longer provides a strong reason for him to believe that the table is really red (Lutz 2018: 1109). After all, his initial justification for his belief that the table is red was based on the original evidence that the table looks red. Hence, it seems that Lucas's initial justification for this belief that the table is red is lost or at least undermined. If he is rational enough, then it seems that he will not continue to hold the belief that the table is red (Lutz 2018: 1109).

4.2.3 Higher-Order Defeaters

Recently, some epistemologists have suggested that there is also a third kind of defeaters – higher-order defeaters – that we should take into account (Christensen 2010; Lasonen-Aarnio 2014; McGrath forthcoming).[84] They aim to introduce this new kind

[84] Some epistemologists also consider the question of how peer disagreement is related to higher-order

of defeaters because the traditional kinds of defeaters (namely rebutting and undercutting ones) are not sufficient to make sense of all kinds of epistemic defeat.

Let us start from an essential feature of the higher-order defeaters, which the previous traditional defeaters seem to lack. This is that, if the justification for my belief is defeated by a higher-order defeater, this seems to entail that I formed my belief via a flawed process. In contrast, in the case of the traditional defeaters, even if the justification for my belief is defeated by them, this in itself need not mean that I formed my belief in a rationally or perceptually defective way (Christensen 2010: 198), nor does it mean that my original evidence was acquired via a flawed process.

To illustrate the previous difference between higher-order and other defeaters, let us return to the previous examples of the traditional defeaters. In the example of a rebutting defeater, the process through which Chris formed her belief that all dogs are black is her ordinary perceptual capacity. This is because she comes to have that belief by seeing only black dogs (on the remote island where she lives). Similarly, in the example of an undercutting defeater, Lucas formed his belief that the table is red by looking at it. Hence, the process through which Lucas formed his belief that the table is red is also his ordinary perceptual capacity.

In both cases, Chris and Lucas formed their initial beliefs by relying on reliable belief-formation processes (i.e., their ordinary perceptual capacities), even if the justifications for their beliefs were then defeated by the traditional defeaters. The fact that the

defeaters. For extended discussions, see Christensen (2010: 186-187) and Kelly (2010: 138-141).

justifications for their beliefs get defeated by the relevant defeaters does nothing to indicate that those processes are not generally reliable. In contrast, it is an essential feature for the higher-order defeaters that, if the justification for one's belief is defeated by a higher-order defeater, then her belief was formed by using a flawed rational or perceptual process.

Let us now imagine a slightly different situation. Imagine that Erica, who is a friend of Lucas, now tells him that, unbeknownst to him, he took a hallucination pill before he looked at the table. Let us also assume that Erica is both reliable and honest. Lucas then comes to have an extra evidence that he took a hallucination pill before he looked at the table. Moreover, the pill has two certain effects. Firstly, he will have a strong hallucination. Secondly, he will not realize that he is in the state of hallucination after taking the pill.

Hence, the extra evidence that Lucas took a hallucination pill can be considered to be higher-order evidence against his initial justification for his previous belief that the table is red. As illustrated by this example, higher-order evidence is not directly relevant to the propositions "that are the subject of the affected belief[s]" (Christensen 2010: 188). The fact that Lucas took a hallucination pill is not directly relevant to the proposition that the table is red. Instead, higher-order evidence is related to an agent's doxastic state when she formed a given belief (Lasonen-Aarnio 2014: 314-315). A higher-order defeater against a belief can be provided by a piece of higher-order evidence if the evidence shows that an agent formed the belief in a way that constituted a rational, a cognitive and/or a perceptual failure (Christensen 2010: 185; Lasonen-Aarnio 2014:

314-316).

In the example of the hallucination inducing drug, the new evidence clearly suggests that Lucas's doxastic state was a result of a flawed process in which the drug played a role when he formed the belief that the table is red. Therefore, this new piece of evidence provides a higher-order defeater against the initial justification that he had for his belief. That is to say, if Lucas becomes aware of the higher-order defeater (i.e., the new evidence), then he also realizes that his belief was never justified in the first place (Lasonen-Aarnio 2014: 317). This is why, in the previous case, the justification for this belief is thus undermined by the higher-order defeater.[85]

4.3 Defeaters and the EDA

We then know that there are three different kinds of defeaters. Nevertheless, in this §4.3, I will argue that the evolutionary debunkers should only argue that the epistemic status of our moral beliefs is undermined only by two kinds of defeaters: the higher-order defeaters that result from their argument and also by the corresponding undercutting defeaters (but only if we understand these defeaters in the way suggested by Scott Sturgeon (2014: 117-118)). I will start from the claim that the fact that our moral beliefs are not tracking the moral truth due to their origin is not a rebutting defeater for these beliefs in 4.3.1. Then, in 4.3.2, I will introduce a view according to which the belief that our moral intuitions are not tracking the moral truth can provide an undercutting

[85] Michael Thune (2010) suggests that higher-order defeaters should be considered to be a kind of partial defeaters. If my original belief is defeated by a partial defeater, it does not necessarily mean that I am thus required to withhold that belief. However, the partial defeater requires me to hold the belief "less firmly" because I should have a lower degree of confidence in this belief due to the defeater (Thune 2010: 358-359).

defeater for our moral beliefs (Kahane 2011: 105-106; Lutz 2018: 1105; Mogensen 2016: 1801; Shafer-Landau 2017: 176-177). For the sake of simplicity, I will call the previous belief (i.e., that our moral intuitions are not tracking the moral truth) NOT-TRACK.

In 4.3.3, I will then introduce an objection to Pollock's account of undercutting defeaters, which was first suggested by Scott Sturgeon (2014). In this sub-section, I will also argue that Sturgeon's own view of undercutting defeaters is right. As a consequence of Sturgeon's objection, the evolutionary debunkers need to insist that there are no evolution-based undercutting defeaters for our moral beliefs unless we have two beliefs – the belief NOT-TRACK and also a higher-order belief about the evolutionary origin of our moral beliefs. Finally, in 4.3.4, I will explain how the evolution-based undercutting defeaters, understood in the way recommended by Sturgeon, can also be translated to corresponding higher-order defeaters for our moral beliefs.

4.3.1 No Rebutting Defeaters for Our Moral Beliefs

Let us start from the rebutting defeaters. Most proponents and opponents of the EDA agree that the fact that our moral beliefs are not tracking the moral truth due to their origin cannot provide a rebutting defeater for our moral beliefs (Clarke-Doane 2017a: 202; Kahane 2011: 106n5; Mogensen 2016: 1801). To illustrate, let's take my moral belief that P as an example. If my argument in Chapters 2 and 3 were along the right lines, this means that I then come to have a new piece of evidence according to which our moral beliefs are not tracking the moral truth. However, it is unlikely that this new

piece of evidence can provide me with a *prima facie* reason for me to believe that ~P. This is because the fact that our moral beliefs are not tracking the moral truth does not specifically indicate whether my moral belief that P is true or false (Mogensen 2016: 1801).

If the previous evidence could really provide a rebutting defeater for my moral belief that P, I should consider giving up my original moral belief that P in order to avoid having inconsistent beliefs. However, there is no inconsistency if I believe that our moral beliefs are not tracking the truth and that P (where P is any moral belief whatsoever). Even if I believed that our moral beliefs are not tracking the moral truth, this belief would thus not give me a reason for me to think that my belief that P is actually false (Clarke-Doane 2017a: 202). My belief that P could still be true, perhaps, by coincidence for example. Hence, the fact that our moral beliefs are not tracking the moral truth due to their origin cannot provide a rebutting defeater for those moral beliefs.

4.3.2 Evolutionarily Origin of Our Moral Beliefs as an Undercutting Defeater

We can then turn to the undercutting defeaters. Many proponents and opponents of the EDA suggest that NOT-TRACK can provide an undercutting defeater for our moral beliefs (Kahane 2011: 105-106; Lutz 2018: 1105; McPherson 2020: 15-19; Mogensen 2016: 1801; Shafer-Landau 2017: 176-177). I will call the view that the NOT-TRACK provides an undercutting defeater for our moral beliefs the Evolutionary Undercutting Account.

In order to make sense of how it could be argued that NOT-TRACK might provide an undercutting defeater for our moral beliefs, we need to start from the way in which our moral beliefs are justified. Let's consider the example of my moral belief that P again. In order to illustrate the way in which my moral belief that P could be claimed to be justified on the basis of evidence, let us consider the following Figure 4.1:

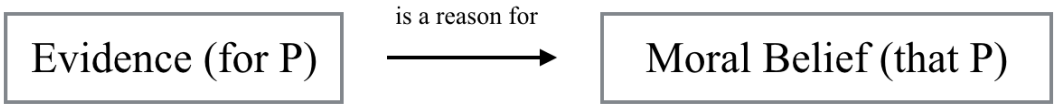

Figure 4.1 – The justification of our moral beliefs

As illustrated by Figure 4.1, when I have a piece of evidence supporting the thought that P, this piece of evidence is a reason for me to believe that P. What kinds of evidence could then provide epistemic reasons that could justify our moral beliefs?

As I mentioned in 3.2.2, many non-naturalist moral realists argue that our moral beliefs are justified by our moral intuitions (Klenk 2017: 785-786; Bedke 2009: 188-189). More precisely, it is often thought that moral intuitions can play an important role in supporting and justifying our moral beliefs because they have certain important distinctive qualities (Kauppinen 2013: 360-361; Lillehammer 2011: 175). In this sub-section 4.3.2, I will explain two such distinctive features: Moral intuitions are thought to be both spontaneous and compelling (Kauppinen 2015: 239-242).[86] According to

[86] According to Antti Kauppinen (2015), moral intuitions have at least five distinctive features (239-242). Apart from being spontaneous and compelling, moral intuitions are also functional, non-doxastic and

Antti Kauppinen (2013), moral intuitions with these two essential features can then play a putative role with regards to how our moral beliefs are justified (361, 365).

Take the moral intuition that that torturing an innocent baby is wrong as an example. This widely shared, strong intuition is presumably for most people both spontaneous and compelling. Firstly, this moral intuition is spontaneous because it is not a result of reasoning or inference from our other moral beliefs. We will have this moral intuition directly once we focus on the moral proposition that torturing an innocent baby is wrong (Kauppinen 2015: 240). Secondly, this intuition is also compelling. That is, the moral proposition that torturing an innocent baby for fun is wrong just very strongly appears to be true to most of us.

We can then return to how moral beliefs can often be justified on the basis of our moral intuitions. Let us consider the moral belief that torturing an innocent baby for fun is wrong. According to moral intuitionists, this moral belief is justified on the basis of our moral intuition that torturing a baby for fun is wrong. Firstly, since that moral intuition is compelling, the moral proposition that torturing an innocent baby for fun is wrong still strongly appear to be true. Furthermore, because that proposition seems to be true to us, we will be strongly inclined to believe that torturing an innocent baby is wrong (Kauppinen 2015: 240). Secondly, that moral intuition is also the basis on which we will form the previous belief. This is because that intuition, because it is spontaneous, is not a result of inference from our other moral beliefs. As a result, our moral intuition

---

they have "a distinctive and diverse phenomenology" that non-moral intuitions lack (Kauppinen 2013: 365).

that torturing a baby is wrong provides a compelling reason for me to believe that torturing a baby is wrong, and thus it can also justify my belief.[87]

Nevertheless, moral intuitionists grant that not all of our moral intuitions with the previous two essential features are able to play an important role in justifying our moral beliefs. Beliefs are justified on the basis of our intuitions only if these intuitions have evidential value. According to a minimal definition of evidential value, an intuition has evidential value only if it has a "positive (not necessarily perfect) correlation with the [relevant] epistemic facts" (Nagel 2013: 181). For example, let us assume that I come to believe that 'Liverpool won the match yesterday' on the basis of rolling a dice. The result of the roll of the dice has no evidential value whatsoever, since it has no correlation with the relevant fact regarding the match yesterday. Hence, it is not reasonable for me to believe that Liverpool did win the match on the basis of that result of roll of the dice, and my belief that 'Liverpool won the match yesterday' is not justified.

Intuitions that are compelling and spontaneous can also lack evidential value. As a result, not all our intuitions with these two essential features can justify our beliefs. We can use the Müller-Lyer illusion as an example, which is illustrated by the following Figure 4.2 (Kauppinen 2015: 239; Ramsey 2019: 79):

[87] Some philosophers instead hold the view that moral intuitions are basically the same kind of mental states as moral beliefs. This alternative view originates from David Lewis (1983: x), who suggests that intuitions and philosophical theories are both opinions. Peter van Inwagen (1997) also claims that our intuitions are simply beliefs (309). See also Williamson (2007: 215). If we accept that the view that intuitions are beliefs, then it would be very easy to explain the tight connection between intuitions and beliefs (Pust 2019: Section 1.1; Ramsey 2019: 79). However, as I mentioned above, moral intuitions seem to have certain distinctive features that moral beliefs generally seem to lack. Therefore, moral intuitions are not commonly considered to be the same kind of mental states as the moral beliefs.

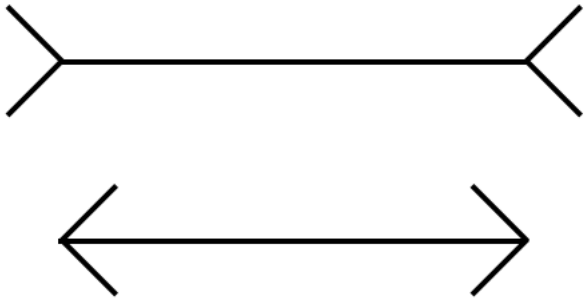

Figure 4.2 – Müller-Lyer illusion

When we look at Figure 4.2, the two lines strongly and immediately appear to differ in length (Ramsey 2019: 79). Our intuition that these two lines are different in length is thus both spontaneous and compelling. Nevertheless, no matter how spontaneous and compelling that intuition is, we should not believe that the two lines are different in length. If we really measure these two lines, we will come to believe that the lines have the same length (Ramsey 2019: 79). This example thus illustrates how intuitions with the two essential features do not necessarily have evidential value (i.e., they do not tend to be correlated with the relevant facts).[88]

Some of our moral intuitions too can be vulnerable to different sources of error, including personal bias, personal interests, misunderstanding, hasty judgments, and so on (Huemer 2005: 137-139; Rawls 1999: 17-18).[89] According to moral intuitionists,

[88] Moreover, even if we firmly believed that these two lines have the same length, this does not also mean that the illusion thus loses its intuitive appeal (Kagan 1989: 15). The two lines still appear to be different in length whenever we look at Figure 4.2. This is because intuitions also seem to be at least in part non-doxastic. See also Kauppinen (2015: 240) and Ramsey (2019: 79).

[89] Michael Huemer (2005) suggests that there are at least 14 kinds of sources of error (137-139).

the moral intuitions that suffer from the previous flaws are unlikely to be correlated with the relevant moral truth and thus lack evidential value. As a result, these moral intuitions, which could well be compelling and spontaneous, fail to be good epistemic reasons that could justify our moral beliefs (Huemer 2005: 105; McMahan 2013: 111).

Hence, moral intuitionists need to suggest a way in which we could "filter out" the moral intuitions that are likely to be in error and thus have no evidential value (Audi (Audi 2015: 129-130; McMahan 2013: 111, 117). In his famous 1951 article "Outline of a Decision Procedure for Ethics", John Rawls introduces a test that can be thought to be filtering out the moral judgments that are likely to be in error (181-183; 1971/1999: 42).[90] He suggests that a competent judge's moral judgment can count as a carefully considered moral conviction only if (Rawls 1951: 182-183):[91]

1. The judge is free from all the foreseeable consequences of holding the judgment. For example, she would not be punished for holding the judgment.

2. The judge's integrity can be maintained when holding the judgment.

3. The judgment is about an actual case, which could well happen in ordinary life, instead of a hypothetical case.

[90] Jeff McMahan (2013) suggests that moral intuitions are a kind of moral judgments (104-105).
[91] According to Rawls (1951), a person can be a competent moral judge only if she (1) is expected to have a certain requisite degree of intelligence; (2) knows relevant non-moral facts; and (3) is reasonable (177-179). He also argues that most, if not all, the competent judges would make identical considered moral judgments of similar cases (Rawls 1951: 184).

4. The judge has carefully inquired into all relevant facts.

5. The judgment is felt to be certain (or in a state of certitude) when the judge holds it.

6. Most competent judges would also have the same judgment on similar cases.

7. The judgment should be intuitive with respect to other ethical principles. Nevertheless, it is not a result from a deliberate application of these principles.

Moral judgments that satisfy the seven criteria above are the carefully considered moral convictions. Of course, there is no guarantee that even the convictions that pass the previous tests must be correlated with the relevant moral truth. Yet, these considered moral judgments are less likely to be subject to the sources of error mentioned above, as they tend to represent the mature convictions that most competent judges would have on similar cases (Rawls 1951: 187). As a result, the carefully considered moral convictions are the moral judgments that are most likely to have evidential value.[92] Hence, a moral belief about a particular case is likely to be justified when it is based on the previous kind of moral convictions (Rawls 1951: 184).[93]

[92] Other intuitionists have suggested slightly different filtering mechanisms. See Mulgan (2006: 2-4) and Sidgwick (1907/1962: 337-343).

[93] This procedure of justifying our moral beliefs and principles can be considered to be a part of the method of "reflective equilibrium", which was first described by John Rawls (1971/1999: 40-45). Reflective equilibrium often means the ideal state in which all of our moral beliefs and principles match our considered moral judgments (Rawls 1971/1999: 18). For extended discussions of the method of reflective equilibrium, see Huemer (2005: 117), Kauppinen (2013: 376-377), Lillehammer (2011: 187-189), and McMahan (2013: 110-112).

We can then return to moral beliefs, undercutting defeaters and the EDA. We can start from considering how an undercutting defeater can undermine the justification of a moral belief. I illustrate how an undercutting defeater works in the following Figure 4.3:

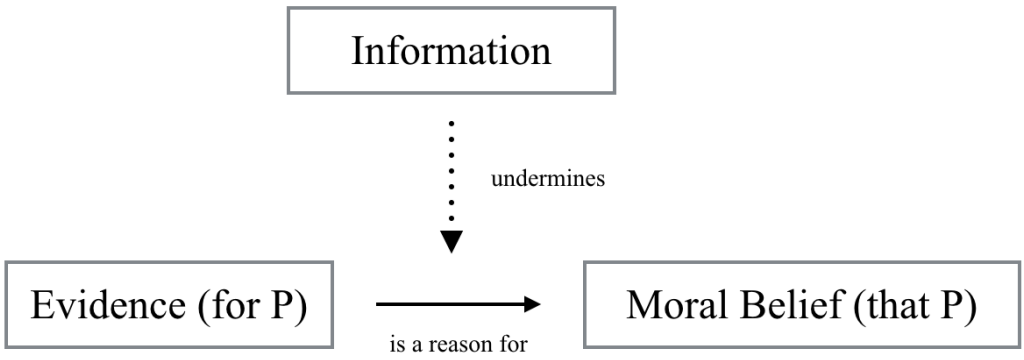

Figure 4.3 – An undercutting defeater

Let's assume that I initially have the moral belief that P, and this is justified on the basis of a piece of evidence that supports the idea that P is true. Let us then stipulate that this piece of evidence is a moral intuition which satisfies all the previous Rawlsian conditions. We can then imagine that I come to have some new information that then undermines the supporting relation between that piece of evidence for P (i.e., the moral intuition that P) and the corresponding moral belief P. In other words, this new information undercuts the way in which the original evidence was a reason for me to believe that P. If this is right, then the previous information is an undercutting defeater for my moral belief that P.

The proponents of the Evolutionary Undercutting Account argue that our understanding

of the evolutionary origin of our moral intuitions is the new piece of information that can undercut the supporting relation between the evidence provided by those intuitions and our corresponding moral beliefs. In Chapters 2 and 3, I discussed the evolutionary origins of our moral beliefs and, as a result, almost all of our moral beliefs do not track the moral truth due to their evolutionary origins. Although I mainly talked about moral beliefs, the very same evolutionary explanation and off-track argument about moral beliefs mentioned in those chapters are also applicable to moral intuition.[94] Hence, in Chapter 2, I already explained the evolutionary origins of our moral intuitions. Also, in Chapter 3, I argued that most, if not all, of our moral intuitions are not tracking the moral truth reliably due to the evolutionary origins of these intuitions.

Based on what I argued in those two chapters, we should then come to have the belief NOT-TRACK. If our moral intuitions are not tracking the moral truth due their evolutionary origins, then these moral intuitions (including even the carefully considered moral convictions, i.e., the intuitions that pass Rawls's tests) fail to be evidence for our moral beliefs. This is because our moral intuitions could be evidence for our corresponding moral beliefs only if their presence increased the likelihood that those beliefs were true. Yet, if our moral intuitions are not tracking the moral truth, then their presence cannot increase the likelihood of our moral beliefs being true. As a result, the previously accepted supporting relation between our moral intuitions and our corresponding moral beliefs is thus undermined by the evolutionary origin of these

[94] For example, in 2.4.2, I explained why our moral beliefs could be thought to be constrained by our emotional systems. In that sub-section, I also explained why our moral beliefs can be considered to be a by-product (i.e., an exaptation) of our emotions. As for the moral intuitions, Antti Kauppinen (2015) suggests that they are "constituted by emotional manifestations of moral sentiments" (237). Arguably, it is plausible to think that the moral intuitions are also thus constrained by our emotional systems and can be considered to be an exaptation as well.

moral intuitions. According to the Evolutionary Undercutting Account, NOT-TRACK should thus be considered to be an undercutting defeater for our moral beliefs.

4.3.3 Sturgeon's Objection to Pollock and Its Consequence for the EDA

In this 4.3.3, I will consider Scott Sturgeon's objection (2014; forthcoming) to John Pollock's view of undercutting defeaters. I will first explain the objection and then its consequences for the EDA.

According to Sturgeon (2014), the satisfaction of Pollock's conditions for what counts as an undercutting defeater is neither necessary nor sufficient for defeat (114-116). Here, however, I will only consider Sturgeon's argument to the conclusion that Pollock's conditions for what counts as an undercutting defeater are not sufficient for defeat. If this is right, it can be argued that there are cases in which there is evidence that satisfies Pollock's definition of undercutting defeat even when no epistemic defeat occurs (McGrath forthcoming). In the case of moral beliefs, this would mean that moral realists might be able to argue that the evidence provided by our moral intuitions is not defeated despite the argument made in the previous sub-section.

To illustrate Sturgeon's objection, let us consider the following example called "Milk Taster":

> Mary is a milk taster. She tastes a bottle of milk to see whether it is spoiled. She comes to have the conclusion that this bottle of milk is okay, and she thinks that

she comes to have this conclusion based solely on taste and not on smell. However, when she tastes the milk, she has a certain complex gustatory and olfactory experience. She is unaware that she is actually basing her conclusion that the milk is okay on smell only and not taste. In other words, unbeknownst to Mary, she actually comes to have the conclusion that the milk is okay merely based on smell. Then, Pierre tells Mary that she is subject to a random olfactory hallucination. Assume that Pierre is honest and reliable. This leads Mary to come to think that her olfactory experience of the milk does not guarantee that the bottle of milk is okay (McGrath forthcoming; Sturgeon 2014: 114-115).

Let us assume that Mary's complex olfactory experience O is evidence for her to have the belief C that the milk is okay (McGrath forthcoming). As a result, Mary is justified in believing that the milk is okay on the basis of O. Let us also stipulate that she in addition believes that she formed that belief C on the basis of taste. Mary then acquires a new belief H that she is subject to a small hallucination (Sturgeon 2014: 115). Belief H then should be considered to be an undercutting defeater according to the Pollock's account of undercutting defeaters. This is because H provides Mary a *prima facie* reason R_H against the view that her original evidence O would support C.

Despite this, it seems that Mary's belief C (i.e., that the milk is okay) is not defeated at all in the previous situation. Let us start from considering how Mary comes to have that belief C. Mary believes that she forms the belief C on the basis of her taste and not on the basis of her smell. Her belief that she formed her belief C on the basis of her taste is false, and yet she still firmly holds this false belief. Now let's assume that she comes

to have the belief H that she has a smell hallucination. Nevertheless, this belief H seems to be irrelevant from her own perspective when she comes to judge whether she should give up her belief C. This is because the belief that she has a smell hallucination provides no reason for her to think that she is mistaken when she forms the belief C.

Therefore, Mary's belief H does not really defeat the supporting relation between her belief C and the basis on which she formed the belief C (i.e., her taste). As long as she holds the belief that she came to have the belief C on the basis of her taste and not on her smell (even if this is a false belief), it is still rational for Mary not to give up C merely because she comes to have the belief H (Sturgeon 2014: 115). Thus, Mary's conclusion C is not defeated. This example then shows that Pollock's account of undercutting defeaters creates too many defeaters. According to Pollock's definition of undercutting defeaters, R_H would be an undercutting defeater for Mary's belief that the milk is okay. However, as I explained, R_H should not be considered to be an undercutting defeater because Mary's belief C is not really defeated.

In order to deal with this problem of Pollock's account of undercutting defeaters, Sturgeon (2014) attempts to revise the definition of undercutting defeaters (117-118). To put it roughly, he suggests that higher-order beliefs are also required for undercutting defeaters to be able to function as defeaters (Sturgeon 2014: 117). Hence, he attempts to revise the definition of undercutting defeaters as follows:

> "Suppose you believe Φ, U is the claim that source S is untrustworthy about Φ, and Φ-BOS is a claim about the basing of your belief in Φ on S…

B(U) undercuts B(Φ) ↔ (∃ x) x = B (Φ-BOS)

[Where ↔ indicates an if-and-only-if relation]" (Sturgeon 2014: 117)

To illustrate Sturgeon's account of undercutting defeaters, let us consider Tim as an example. Imagine that Tim has the belief Φ that 357 is a prime number, but he came to have this B (Φ) on the basis of rolling a dice. Now Tim also acquires another belief U that throwing dice is an untrustworthy method to form beliefs about prime number. According to Sturgeon's view of undercutting defeaters, B(U) undercuts B(Φ) if and only if Tim also has the belief Φ-BOS that he formed the B(Φ) on the basis a dice roll. For the sake of simplicity, I will call the method of throwing dice to form a belief about prime numbers 'S'.

Let us start from considering why it is necessary for Tim to have B(Φ-BOS) in order to have an undercutting defeater for B(Φ). This is because B(U) itself is not enough to undercut B(Φ) unless Tim also has the belief about the origin of B(Φ). B(U) is merely a belief about the untrustworthiness of S. In order for B(U) to undercut B(Φ), it is also required that Tim furthermore has the belief that the origin of B(Φ) is really S. B(Φ-BOS) is thus necessary for B(U) to undercut B(Φ).

We can then consider why it is also sufficient for Tim to have an undercutting defeater for B(Φ) if Tim has B(Φ-BOS) and also B(U). Let's assume that Tim now has two beliefs: B(Φ-BOS) and B(U). B(Φ-BOS) is a belief about the origin of B(Φ) (i.e., S), and B(U) is a belief about the untrustworthiness of S. These two beliefs thus jointly

entail that the origin of B(Φ) is untrustworthy, and consequently B(Φ) is defeated. As a result, it is sufficient for Tim to have the undercutting defeater for B(Φ) if he has B(Φ-BOS) and also B(U).

Hence, in order for Tim to have an undercutting defeater for B(Φ), it is necessary and sufficient that he has both a higher-order belief Φ-BOS about the source of B(Φ) and the belief (U) of that source being untrustworthy. The alleged undercutting defeaters of Pollock's account of undercutting defeaters, however, clearly do not require any higher-order beliefs such as B(Φ-BOS) as we saw in the case above.

Let us then consider the Milk Taster example again. Following what Sturgeon argued, Mary's belief H (i.e., she has a smell hallucination) itself is not sufficient to undercut her belief C (i.e., the milk is okay). In other words, when there is an undercutting defeater for her belief C, she also needs to have a higher-order belief about the way in which she formed the belief that C, in addition to her belief H (Sturgeon 2014: 117). This means that Mary has an undercutting defeater for her belief C if and only if she believes that (1) she has a smell hallucination and also that (2) her olfactory faculty is the basis of her belief C that the milk is okay.

We then need to consider the consequences of Sturgeon's view on undercutting defeaters for the EDA. In my mind, this is where the evolutionary debunkers need to reconsider what the relevant undercutting defeaters for our moral beliefs that follow from their views actually are. In the example above, Mary's belief C would, on Sturgeon's view, be an undercutting defeater if and only if Mary has the belief about

the origin of her belief C and the belief that this origin is flawed. Following the same line of reasoning, we would have an undercutting defeater for our moral beliefs if and only if we also had certain beliefs about the origin of our moral beliefs and also certain beliefs of that origin being flawed. In other words, NOT-TRACK itself is also not sufficient to be an undercutting defeater for our moral beliefs.

Thus, evolutionary debunkers should claim that there is an undercutting defeater for our moral beliefs only when we believe both NOT-TRACK and that the origin of our moral beliefs can be wholly explained by natural selection. I will call the evolutionary explanation of the origin of our moral beliefs EVO-ORIGIN. Let us then consider the following Figure 4.4, which illustrates the basic crux of how our beliefs EVO-ORIGIN and NOT-TRACK jointly provide an undercutting defeater for our moral beliefs:

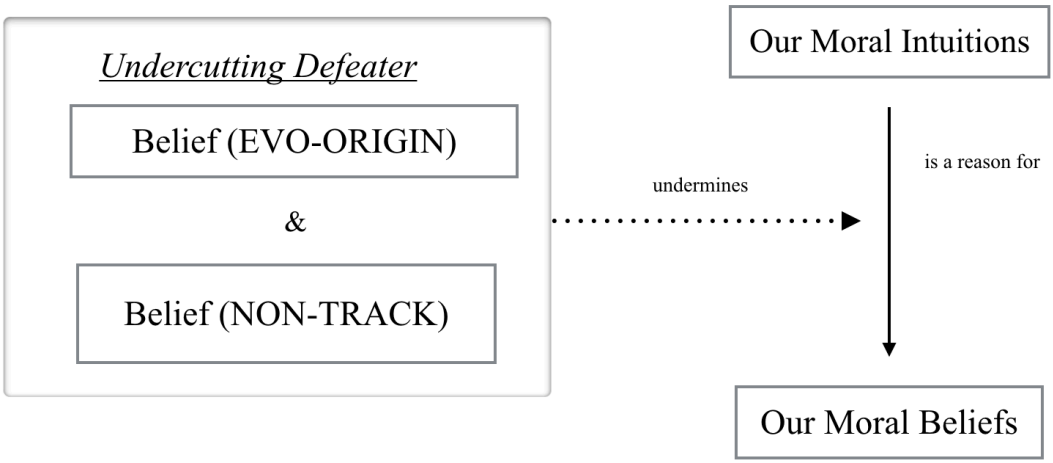

Figure 4.4 – EVO-ORIGIN and NOT-TRACK as an undercutting defeater

According to Sturgeon's account of undercutting defeaters, our belief NON-TRACK is

required even if it is not sufficient to be the undercutting defeater for our moral beliefs on its own. This is because a higher-order belief about the way in which we form our moral beliefs is also required, in addition to our belief NOT-TRACK.

Moral realists may argue that they do not believe in EVO-ORIGIN and NOT-TRACK, and so there aren't any undercutting defeaters at least for their moral beliefs. However, in Chapter 2 and Chapter 3, I have already provided compelling arguments and evidence for why we should believe both EVO-ORIGIN and NOT-TRACK. This means that at least all rational persons who are aware of those arguments and who form their beliefs on the basis of the most compelling arguments (including many of those who are currently realists) also have an undercutting defeater for their moral beliefs.[95]

To sum up: I argued that Sturgeon's objection to Pollock's account of undercutting defeaters is correct. Hence, in order to provide an undercutting defeater for our moral beliefs, we must have both beliefs EVO-ORIGIN and NOT-TRACK. As I explained in chapters 2 and 3, we should believe both EVO-ORIGIN and NOT-TRACK. Therefore, I argue that there is an undercutting defeater for our moral beliefs because of the EDA. The epistemic status of our moral beliefs is thus undermined due to this undercutting defeater.

[95] As I mentioned in 4.3.2, some moral realists, who are also moral intuitionists, may argue that our moral beliefs are justified on the basis of our carefully considered moral convictions. However, it is generally believed that only the considered moral convictions of competent judges can play a justificatory role. In this situation, it is unlikely that any rational and competent judges would ignore the relevant empirical evidence and arguments. In Chapters 2 and 3, I already provided compelling arguments for why we should believe both EVO-ORIGIN and NOT-TRACK. As a result, I believe that all rational and competent judges would also believe EVO-ORIGIN and NOT-TRACK and thus also accept that there is an undercutting defeater for our moral beliefs.

4.3.4 Higher-Order Defeaters for Our Moral Beliefs

In the previous section 4.3.3, I argued that there is an undercutting defeater for our moral beliefs if we understand undercutting defeaters in the way suggested by Sturgeon. However, at this point you may notice that higher-order defeaters, which I already explained in 4.2.3, are very similar to Sturgeon's undercutting defeaters. Both kinds of defeaters for our beliefs now require higher-order belief about the source of those beliefs. In this 4.3.4, I will argue that the EDA-based undercutting defeaters can also be translated to corresponding higher-order defeaters.

We can start from how Sturgeon's undercutting defeaters can be generally translated into higher-order defeaters (McGrath forthcoming). Let us use Tim's belief Φ that 357 is a prime number as an example again. As we saw above, there is a Sturgeonian undercutting defeater for B(Φ) if and only if Tim has two beliefs: (1) belief Φ-BOS (i.e., the source of Tim's belief Φ is S) and (2) belief U (i.e., S is untrustworthy). B(Φ-BOS) and B(U) jointly imply that the source of B(Φ) is untrustworthy, and thus there is an undercutting defeater for B(Φ).

However, in the previous case, Tim also has a higher-order defeater for his belief Φ. Following what I explained in 4.2.3, there is a higher-order defeater for B(Φ) if and only if Tim has a belief about the origin of his belief and a belief of that source being unreliable. The belief about the origin of his belief is the B(Φ-BOS), and the belief of that source being unreliable is B(U), where these are the beliefs that were above required for having the undercutting defeater too. Hence, the conjunction of B(Φ-BOS) and B(U) does not merely provide an undercutting defeater for B(Φ), but it also

provides a higher-order defeater for B(Φ) as well.[96]

Let us then consider how the EDA-based Sturgeonian undercutting defeaters can be translated to corresponding higher-order defeaters too. As I explained in 4.3.3, there is a Sturgeonian undercutting defeater for our moral beliefs when we believe both NOT-TRACK and EVO-ORIGIN. The translation is as follows: If we believe EVO-ORIGIN, then we have a higher-order belief that our moral beliefs are formed via a certain evolutionarily influenced process. And, if we also believe NOT-TRACK, then we have a belief that this process is an unreliable mechanism.

Thus, if we believe NOT-TRACK and EVO-ORIGIN, the conjunction of these two beliefs is that our moral beliefs are formed by an unreliable mechanism. As I explained in 4.2.3, there is a higher-order defeater for our moral beliefs when we have a higher-order belief that our moral beliefs are formed via an unreliable process. As a result, the conjunction of our beliefs NOT-TRACK and EVO-ORIGIN provides us with a higher-order defeater for our moral beliefs, and so the relevant EDA-based undercutting defeater can be translated into the higher-order defeater for our moral beliefs. And, the epistemic status of our moral beliefs is thus undermined because of this higher-order defeater too.

[96] Matthew McGrath (forthcoming) acknowledges that, if the undercutting defeaters are inferentially acquired beliefs, then they cannot be translated into higher-order defeaters. However, as mentioned above, our moral beliefs are generally thought to be formed on the basis of our moral intuitions, and therefore most of our moral beliefs are not inferentially acquired. Let's consider the famous case in which I see a hoodlum burning a cat as an example (Harman 1988: 120). In this case, I first have the moral intuition that her act is wrong, and I then come to have the moral belief that burning a cat is wrong on the basis of that moral intuition. However, importantly I do not infer that moral belief from my intuition that burning a cat is wrong and so this type of typically moral beliefs are not inferentially acquired. This is why at least most undercutting defeaters can be translated into higher-order defeaters in the case of moral beliefs.

Let me then summarize what I argued in §4.3. I first argued that there are no rebutting defeaters for our moral beliefs in the premises of the EDA. I then introduced Sturgeon's objection to Pollock's account of undercutting defeaters. I argued that, as a consequence of Sturgeon's objection, evolutionary debunkers need to insist further that we should believe both EVO-ORIGIN and NOT-TRACK. If we have both of these beliefs, EVO-ORIGIN and NOT-TRACK, these two beliefs will jointly provide an undercutting defeater for our moral beliefs. Finally, I also argued a higher-order defeater for our moral beliefs can also be constructed on the basis of the first premises of the EDA. This is because evolutionary debunkers can translate the undercutting defeater based on their views into a higher-order defeater for our moral beliefs. As a result, the epistemic status of our moral beliefs is undermined due to an undercutting defeater that is also at the same time a higher-order defeater.

4.4 Externalists and the EDA

In the previous section, I argued that the epistemic status of our moral beliefs is undermined because we are aware of a higher-order defeater and an undercutting defeater for these beliefs. However, there is a debate about justification in epistemology concerning whether justification is internal or external. The internalists in this debate argue that the factors which make an agent's moral beliefs justified or unjustified must be available to her and so they must be internal to her psychological make-up. In contrast, the externalists deny this – they think that such factors need not always be available to the agent and so they can be, for example, worldly facts too. These factors are considered to be external because "an agent may be warranted in her beliefs even if

she cannot recognize that she is" (Shafer-Landau 2003: 243).[97]

So far in this thesis, I have assumed that internalism is right. In other words, I have assumed that the factors which make our moral beliefs justified or unjustified must be available to us. In this Chapter 4, I have also followed the mainstream epistemic epistemologists' views on defeaters. That is, I have understood them in an epistemic internalist way (Greco 2010: 156-159; Janvid 2017: 702; Vahid 2011: 151-152).[98] However, at this point, moral realists could make the following objection. They could reject the internalist assumptions that I have relied on and suggest that what justifies our moral beliefs is not something internal to our psychological and cognitive perspectives. As a result, they could argue that any beliefs about the evolutionary origin of our moral beliefs and the untrustworthiness of this origin just cannot undermine the external justification that we have for our moral beliefs. The epistemic status of our moral beliefs is thus not undermined.

Could the realists avoid the challenge posed by the EDA in this way? I think the answer is no, or so I will argue next. But I want to make two remarks first. Firstly, I will remain neutral between whether internalism or externalism is true. Hence, in this section, I will argue that, even if we assumed that externalism were true, this would not help the moral realists to avoid the EDA challenge.

---

[97] Epistemic externalism is sometimes defined as the view that rejects the KK-Principle in epistemology (Okasha 2013: 80; Schantz 2004: 9). According to the KK-Principle, if a subject knows a proposition P, then she must grasp the proposition that she knows that P and must be in a position to know that she knows that P (McHugh 2010: 231; Okasha 2013: 80). However, the externalist view that rejects the KK-Principle is better thought of an account of knowledge rather than as an account of justification. In this thesis, I mainly focus on the justification of our moral beliefs instead of moral knowledge. It is therefore beyond the scope of this thesis to discuss that kind of externalist theory of knowledge.

[98] For a discussion of how to understand defeaters in an externalist way, see Janvid (2017).

Secondly, reliabilism is often considered to be the dominant externalist theory of justification in epistemology (Okasha 2013: 80). According to reliabilism, if an agent's beliefs are produced by a belief-formation process that is likely to produce more true beliefs than false beliefs, then her belief-forming process is reliable (Shafer-Landau 2003: 273).[99] Let us then imagine that an agent now has the belief that P. Reliabilists suggest that the agent's belief that P is justified if and only if her that belief was formed by relying on a reliable process. Moreover, even if the agent knows nothing about the reliability of that belief-forming process, her belief that P will still be justified. This is why reliabilism is an externalist theory of justification. Even though there are other forms of externalism, for the sake of simplicity, I will only focus on the reliabilism in here.[100]

How could the realists formulate a response to the EDA on the basis of the previous kind of externalist reliabilism?[101] According to reliabilism, our moral beliefs are justified when these beliefs are formed by a reliable belief-formation process. Let us assume that our moral beliefs are all formed by a reliable belief-formation process. We then also acquire a belief about the evolutionary origin of our moral beliefs (i.e., belief EVO-ORIGIN) and a belief about that origin being unreliable (i.e., belief NOT-TRACK). At this point, the reliabilists could insist that our beliefs EVO-ORIGIN and NOT-TRACK could not really change the reliability of our moral belief-formation

[99] Recently, some epistemologists have also explored other ways of understanding reliability in terms of modal conditions. For example, I introduced the modal reading of reliability in Chapter 3.
[100] For instance, in 3.2.1, I discussed the causal theory, which is also an externalist theory.
[101] For an example of moral realist who also is a reliabilist, see Shafer-Landau (2003: 272-275).

process. Our moral beliefs are still produced by a reliable process that is likely to produce more true beliefs than false beliefs. As a result, they would remain justified even if we (perhaps falsely) believed that our moral beliefs were produced by an unreliable process.

Unfortunately, this response to the EDA does not work. This is because the externalist moral realists' previous assumption that our moral beliefs are all formed by a reliable belief-formation process is actually mistaken. As I already argued in Chapters 2 and 3, the evolutionary origin of our moral beliefs is an unreliable belief-formation mechanism – it just does not even track the moral truth. This means that our moral beliefs have always been formed by an unreliable mechanism. If we then assume that externalist reliabilism is true, all this assumption really entails is that our moral beliefs would have never been justified in the first place, and so the realists would have just mistakenly thought that our moral beliefs were once justified. Even if the externalist moral realists were right to argue that our beliefs EVO-ORIGIN and NOT-TRACK could not change the reliability of the processes through which we came to have our moral beliefs, this would not entail that our moral beliefs would be justified given their evolutionary origin.[102]

The evolutionary debunkers thus need to have two different arguments against the

[102] David Enoch (2011), who is a robust moral non-naturalistic realist, has put forward a similar argument. He suggests that we cannot "justifiably form a belief using what [we] know is an unreliable method" (Enoch 2011: 161). If we are aware of the fact that our moral beliefs are formed in an unreliable way, then there is an internal defeater for our moral beliefs. This is because the unreliability of the belief-formation method would "defeat whatever initial justification [we] may have had for [our moral beliefs]" (Enoch 2011: 161). In order to reject the EDA, Enoch (2011) instead argues that we should not believe NOT-TRACK even if we have the belief EVO-ORIGIN. I will discuss Enoch's objection to the EDA in Chapter 6.

internalist and the externalist realists. In order to pose a challenge to the internalist realists, the debunkers need to argue that beliefs EVO-ORIGIN and NOT-TRACK jointly provide us with a higher-order defeater and also an undercutting defeater for our moral beliefs. In addition, if the debunkers want to pose an EDA-based challenge to the externalist realists, then they should also argue that, if we assume that externalism is true, then there never was a justification for our moral beliefs in the first place. This is because our moral beliefs were formed by relying on an unreliable mechanism (i.e., they have an evolutionary origin).

4.5 Conclusion

In this conclusion, I will first summarize what I argued in this Chapter 4, and after this, I will also summarize the whole EDA as it has emerged from the previous 3 chapters of this thesis. In this Chapter 4, I first introduced and evaluated three different kinds of defeaters – rebutting, undercutting and higher-order defeaters. I argued that the fact that our moral beliefs are not tracking the moral truth due to their evolutionary origin fails to provide us with a rebutting defeater for these moral beliefs.

I then introduced an objection to John Pollock's account of undercutting defeaters and also argued, based on Scott Sturgeon's Milk Taster objection, that undercutting defeaters require higher-order beliefs. As a result, I suggested that our beliefs NOT-TRACK and EVO-ORIGIN jointly provide an undercutting defeater for our moral beliefs. I furthermore suggested that evolutionary debunkers can also translate this EDA-based undercutting defeater into a higher-order defeater.

Finally, in the end of this chapter 4, I explained how realists could try to respond to the EDA by relying on externalism about justification. Nevertheless, I also argued that, if we assume that externalism is true, then our moral beliefs would have never been justified in the first place given their unreliable evolutionary origin. Based on what I have argued in this chapter, I will then conclude that the epistemic status of our moral beliefs is clearly undermined, if we believe both NOT-TRACK and EVO-ORIGIN (as we should due to the arguments of Chapter 2 and 3).

Let me then summarize the EDA in the form that I have defended it so far in Part 1 of this thesis. In Chapter 2, I described how evolution has significantly influenced and shaped our moral beliefs and, as a result, the evolutionary debunkers can wholly explain what the moral beliefs we have come to have on the basis of evolution. In Chapter 3, I argued that, if our moral beliefs have that kind of an evolutionary origin, then our moral beliefs are not tracking the moral truth. Finally, in this chapter 4, I suggested that, if we are aware of the previous two conclusions of my arguments, then the epistemic status of our moral beliefs will be undermined. In conclusion, based on what I have argued in Chapters 2-4, I conclude that the EDA is a sound argument and that the epistemic status of our moral beliefs is undermined accordingly.

In the two chapters below, Chapters 5 and 6, I will then introduce two of the strongest objections to the EDA: the Conceptual Truth Objection and the Third-Factor Objection. In these chapters, I will also argue that both of these objections are mistaken, and so they give us no reason to reject the conclusion of the EDA.

Part 2

Objections to Evolutionary Debunking Argument and Replies

Chapter 5

The Conceptual Truth Objection

5.1 Introduction

In this Chapter 5, I will focus on the first of the two stronger realist objections to the EDA. I will call this objection the Conceptual Truth Objection. As explained in Chapter 2, in order to create an epistemological challenge for the moral realists, all evolutionary debunkers begin from explaining the evolutionary origin of our moral beliefs. The Conceptual Truth Objection, however, aims to argue that the epistemic status of our moral beliefs would not be undermined even if the debunkers could wholly explain the origin of these beliefs in evolutionary terms.

The moral realists who could pursue this line of response include both analytic naturalists (Finlay 2014; Jackson 1998; Smith 1994) and realists who defend the idea of "moral fixed points" (Cuneo & Shafer-Landau 2014).[103] I will call these versions of moral realism according to which the moral truths are conceptual truths forms of 'conceptual truth realism'. Unlike other sorts of moral realism, conceptual truth realism suggests that our moral beliefs can be justified by conceptually analyzing our normative terms and by finding moral propositions that are conceptual moral truths. According to the conceptual truth realists, the evolutionary forces are then unlikely to have any debunking influence on this way in which we are able to justify our moral beliefs, and

[103] Cuneo and Shafer-Landau (2014) are the only conceptual truth realists who have explicitly pursued this line of a response to the EDA in their works.

therefore the evolutionary origins of our moral beliefs are insufficient in themselves to undermine the epistemic status of those beliefs. If they were right, then the EDA would fail to pose a real epistemological challenge to the conceptual truth realism.

The main objective of this Chapter 5 is to defend the EDA against the Conceptual Truth Objection. This chapter has four parts. In §5.2, I will begin by introducing the Conceptual Truth Objection and how it could be used to argue that the EDA fails to pose an epistemological challenge to conceptual truth realism. I will then outline and evaluate three potential ways in which the evolutionary debunkers could try to reject the previous objection. Firstly, in §5.3, I will introduce Evers and Streumer's objection to the moral fixed points theory (2016), but I will also argue that this objection cannot help the evolutionary debunkers to reject the Conceptual Truth Objection. Then, in §5.4, I will introduce an R.M. Hare-styled argument against the application of the classical theory of concepts to moral concepts on which the Conceptual Truth Objection is based (1952/1991). In that sub-section, I will argue that the classical theory of moral concepts, which is adopted by the conceptual truth realists, is mistaken. Finally, in §5.5, I will focus on the conceptual truth realists' view of our competency for justifying our moral beliefs by doing conceptual analysis. I will argue that we have not evolved to have a competency to discover which of our moral beliefs are true by doing conceptual analysis, even if there were such conceptual moral truths.

5.2 The Objection

In this §5.2, I will introduce the Conceptual Truth Objection to the EDA. As mentioned above, some moral realists suggest that moral truths are conceptual truths. According

to these 'conceptual truth realists', we can also know these conceptual moral truths in a specific way, that is, by doing conceptual analysis. Furthermore, the conceptual truth realists argue that the fact that we are able to do conceptual analysis reliably also explains how we can access the conceptual moral truths.

The essential idea of the Conceptual Truth Objection is then to suggest that the EDA cannot undermine the epistemic status of the moral beliefs, provided that we are able to justify these moral beliefs by doing conceptual analysis. Indeed, the defenders of the objection in question can also accept the evolutionary origin of our moral beliefs. Nevertheless, they attempt to argue that the fact that the evolutionary forces have influenced our moral beliefs need not itself necessarily mean that we are unable to justify our moral beliefs by doing conceptual analysis. If the defenders of the objection were right, then the epistemic status of our moral beliefs is thus protected against the EDA.

To further illustrate how the Conceptual Truth Objection works, we can use the following example of unreliable testimony. In our daily life, we come to have many true beliefs based on unreliable testimony. Consider a situation in which a layman tells me that the equation '$E=mc^2$' is true and as a consequence I come to believe that the equation is true. Obviously, the equation is really true. However, the layman does not know why it is true. He does not understand what the equation says, and he might have even heard the equation from a random TV show.[104] In this situation, although I have

[104] For example, in an episode of the American situation comedy "The Big Bang Theory", one of the character – Penny – mentioned the equation '$E=mc^2$' even if she also conceded that she just remembered the equation and had "no idea [of] what it means" (Lorre et al. 2007).

come to have a true belief based on unreliable testimony, this previous belief can still be justified in other ways. For instance, I can go to read physics textbooks or ask a professor of physics. If I do so, the troublesome origin of my belief will not be able to undermine the justification I will have for the belief in question.

By relying on a similar line of reasoning, the defenders of the Conceptual Truth Objection argue that our moral beliefs can still be justified in a reliable way even though these beliefs could have an unreliable origin. In other words, the evolutionary origin of these beliefs cannot undermine the epistemic status of our moral beliefs as long as we can acquire new justification for them by doing conceptual analysis. The Conceptual Truth Objection can thus be formulated roughly in the following way:

The Conceptual Truth Objection Argument

1. Moral truths are conceptual truths.

2. If the moral truths are conceptual truths, then we are able to discover which moral propositions are true by doing conceptual analysis.

3. Even if the evolutionary forces have significantly influenced and shaped our moral beliefs, we still have the competency to discover which of our moral beliefs are true by doing conceptual analysis.

4. If we can come to discover which of our moral beliefs are true by doing conceptual analysis, our moral beliefs can be understood to track the moral truth.

5. Conclusion: The epistemic status of these moral beliefs is thereby not undermined.

In the rest of this §5.2, I will explain the premises of the Conceptual Truth Objection Argument. Firstly, 5.2.1 will explain the 1st premise and also outline why the conceptual truth realists think that the moral truths are conceptual truths. Then, 5.2.2 will explain the 2nd and 3rd premises. In that sub-section, I will introduce how the conceptual truth realists think that we, as competent speakers, are able to know the conceptual moral truths by doing conceptual analysis. Finally, 5.2.3 will explain the 4th premise and the conclusion of the argument. I will then explain why the defenders of the objection think that the epistemic status of our moral beliefs is thereby not undermined by their evolutionary origin as long as these moral beliefs are justified via conceptual analysis.

5.2.1 Conceptual Moral Truths

Conceptual truth realists take the moral truths to be conceptual truths. But, what are conceptual truths? To explain what conceptual truths are, let us first consider what concepts are. The conceptual truth realists seem to have adopted the classical theory of the nature of concepts.[105] According to this theory, concepts have three essential features.[106] Firstly, concepts have a reductive structure. That is, we can analyze a

[105] The classical theory of the nature of concepts is famously attributed to G.E. Moore (1899), and Bertrand Russell (1903/2010). For an overview of the more recent development of the theory, see Finlay (2014).

[106] According to the classical theory of the nature of concepts, concepts also have other important features than the three features discussed here. For example, concepts have a referential function of picking out things at the metaphysical level (Laskowski & Finlay 2017: 537). In this thesis, I will not explain all features of concepts as some of them are not directly relevant for the Conceptual Truth Objection. For an overview of all features of concepts according to the classical theory of the nature of

concept and reduce it to its constituent parts (Laskowski & Finlay 2017: 538). Secondly, concepts have a descriptive structure. The reductive structure of concepts enables us to define them in descriptive terms. To explain these two features of concepts, we can consider the concept of <vixen> as an example. By analyzing the concept <vixen>, we can reduce it to simpler constituent parts that can be given in terms of the concept's descriptive definition. The descriptive definition of <vixen> is 'female fox' which reveals that the constituent parts of the concept <vixen> are the descriptions '… is female' and '… is a fox'.

Finally, the third important element of the classical theory of concepts is that, as competent speakers, we possess a concept if and only if we can psychologically grasp its descriptive definition (Laskowski & Finlay 2017: 538). This psychologically grasp can be argued to consist of the fact that speakers must be able to apply the concept on the basis of the descriptive definition of the concept. Hence, the descriptive definition of the concept can also be considered to be functioning as the implicit principle that guides our application of the concept insofar as we master the concept. Let us consider the concept <vixen> as an example again. A competent speaker who possesses the concept <vixen> must be able to psychologically grasp the descriptive definition of the concept <vixen> as well. That is to say, when she applies the concept in question, she must be guided by an implicit principle that corresponds to the descriptive definition of the concept <vixen>. This also means that she is able to use the concept with competence by applying it to objects that are 'female foxes'.[107]

concepts, see Cuneo and Shafer-Landau (2014: 409-412) and Laskowski and Finlay (2017).

[107] In this thesis, I rely on a modest requirement for a speaker to be able to apply a concept with competence. Some have also defended stronger requirements for a speaker to be considered to be a

According to the classical theory of concepts, there will be conceptual truths given that, concepts have reductive and descriptive structures. On this view, conceptual truths are propositions that are true in virtue of their constituent concepts (Cuneo & Shafer-Landau 2014: 410-411; Moore 1899: 180). Let us consider the proposition 'all vixens are female foxes', which is a classic example of conceptual truths, as an example. This proposition is true in virtue of its constituent concepts. The concept <vixen> means 'female fox' because the latter is the former's descriptive definition. We can then start from the trivially true claim 'all female foxes are female foxes'. Because the concept <vixen> just means 'a female fox', we are allowed to substitute this concept for the first instantiation of the description 'female foxes' in the previous trivial truth. If we do so, we get the proposition 'all vixens are female foxes'. As a consequence, this proposition is not only true, but it is also necessarily, conceptually true.[108]

All conceptual truth realists share the view that at least some moral truths are conceptual truths. Furthermore, they all think that some moral beliefs and propositions are conceptually true in a similar way in which the proposition 'all vixens are female foxes'

competent user of concept. According to one strong requirement, for example, a competent speaker must be able to associate the concept with its descriptive definition in her mind explicitly (Huemer 2015: 53-54). Very often, however, we seem to be able to use some concepts with competence without realizing their relevant descriptive definitions in explicit forms (Suikkanen 2016: 359-360). For instance, we are able to apply the concept <book> correctly, but it is very rare that we have the descriptive definition of the concept <book> explicitly in our mind when we use the concept. In addition, the modest requirement is also compatible with the strong requirement but not vice versa. If a speaker can have the relevant descriptive definition of a concept in mind explicitly, she must be able to apply the concept correctly as well but not vice versa. Hence, hereafter I will assume that the conceptual truth realists hold the modest requirement for a speaker to apply a concept with competence.

[108] Sometimes conceptual truths are also thought to be the same as the 'analytic truths'. For example, the proposition that 'all vixens are female foxes' is a sentence that is analytically true. This means that this proposition is true merely in virtue of the meaning of its constituent words (Ayer 1936/1946: 16; Landini 2011: 223). In this thesis, I do not attempt to explicitly distinguish between conceptual truths and analytic truths. See also Boghossian (1994; 1997) and Daly (2010: 45-46).

is true. But, why do conceptual truth realists think that the moral truths are conceptual truths? Different versions of conceptual truth realism differ in here. I will next consider two representative versions of conceptual truth realism: (i) analytic naturalism (Jackson 1998) and (ii) a realist account based on "moral fixed points" (Cuneo & Shafer-Landau 2014).

(i) Analytic naturalism

Analytic naturalism can be thought to be a version of conceptual truth realism. This is because analytic naturalists aim at providing reductive accounts of moral concepts through conceptual analysis. Because they believe that moral concepts have reductive and descriptive structures, they attempt to reduce moral concepts like <right>, <wrong>, <good> or <bad> to their alleged naturalistic and descriptive constituents. Let us consider the moral concept <right> and also assume that X is a naturalistic description. Analytic naturalists aim to reduce the moral concept <right> to the naturalistic description X by trying to identify the correct reductive definition of the following form: To be right is by definition to be X (Smith 1994: 36).[109] For example, some utilitarians have thought that the moral concept <right> can be defined with the naturalistic description 'maximizes happiness'. In other words, on this view, when we say that an act is right, what we mean is that this act maximizes happiness.[110]

Furthermore, according to analytic naturalism, a speaker who can apply and use a moral

[109] Michael Smith (1994) also calls this form of analytic naturalism "(non-subjective) definitional naturalism" (35).
[110] Not all utilitarians are moral cognitivists, let alone analytic naturalist realists. For example, J.J.C. Smart (1973) is a utilitarian who also is a moral non-cognitivist.

concept with competence must also be able to psychologically grasp the relevant naturalistic descriptive definition of that moral concept. Consider the moral concept <right> as an example again. According to the view under investigation, a competent user of this concept must be able to psychologically grasp the relevant naturalist definition of this concept, where this psychological grasp consists of the fact that the previous definition's constituent descriptions function as the implicit principles that guide her application of the moral concept <right>. For example, some utilitarians have thought that a competent speaker thus applies, in both her speech and thought, the moral concept <right> to actions that satisfy the naturalistic descriptive definition 'maximizes happiness'.

Analytic naturalism holds that at least some moral truths are conceptual truths. As mentioned above, conceptual truths are propositions that are true in virtue of their constituent concepts. Let us consider the moral proposition 'it is right to maximize happiness' as an example. We can assume, at least for the sake of an argument, that the moral concept <right> really means 'maximizes happiness'. In this situation, the proposition 'it is right to maximize happiness' would be true because the moral concept <right> is analytically equivalent to its descriptive definition 'maximizes happiness' (Toppinen 2014: 16). The proposition 'it is right to maximize happiness' would thus be true exactly in the same way as the proposition 'actions that maximize happiness maximize happiness' is true because these just are the one and the same proposition. Thus, we can get the former proposition from the latter by substituting a concept in the place of its constituent descriptions. This is why the analytic naturalists think that moral propositions of this kind are conceptual truths – they are true in virtue of their

constituent concepts.

(ii) The realist account of "moral fixed points"

Cuneo and Shafer-Landau's realist account of "moral fixed points" (2014) is thought to be another version of conceptual truth realism. The moral fixed points theory too suggests that moral concepts have reductive and descriptive structures, at least partially. Unlike analytic naturalism, however, this theory does not attempt to reduce the moral concepts like <right> or <wrong> into naturalistic descriptions like 'maximizes happiness'. According to this theory, rather, moral concepts are not identical with or equivalent to any naturalistic descriptions, even though some relevant descriptive definitions are contained in the structure of the moral concepts (Cuneo & Shafer-Landau 2014: 403n9). Cuneo and Shafer-Landau thus suggest that we, as speakers, can follow the descriptive elements of the definitions of moral concepts that guide our application of the moral concepts if we can know the essences of the moral concepts.

On this view, however, what are the essences of the moral concepts? We can start from considering the thick concepts like <selfish> and <honest>. We know how to apply these thick concepts because the conditions under which acts qualify as selfish or honest are already fixed by their descriptive contents (Cuneo & Shafer-Landau 2014: 406). The essential idea of the moral fixed points theory is that the thin moral concepts like <right> or <wrong> "behave much more like thick ones than many have believed" (Cuneo & Shafer-Landau 2014: 406).[111] That is, the conditions under which acts

[111] Moral concepts are considered to be 'thick' (and thus not 'thin') if they mix together in some way the evaluative and non-evaluative descriptions (Väyrynen 2013: 1-3). For an overview of how philosophers have attempted to draw a distinction between thick and thin concepts, see Väyrynen (2013: 1-7).

qualify as right or wrong likewise are fixed by the descriptive elements of the essences of those moral concepts in question.

Consider the descriptive concept of <killing a child merely for fun> as an example. According to the moral fixed points theory, this concept belongs to the essence of the moral concept <wrong>. Necessarily, anything that falls under the concept of <killing a child merely for fun> also falls under the moral concept of <wrong> (Cuneo & Shafer-Landau 2014: 413). Likewise, a competent speaker who possesses the moral concept <wrong> can apply that concept to wrong acts of killing a child merely for fun. This is because she is aware of the descriptive concept of <killing a child merely for fun> being a part of the concept of <wrong> (Cuneo & Shafer-Landau 2014: 410).

The defenders of the moral fixed points theory then suggest that there are many moral propositions that are also conceptual truths, and they also call these propositions 'moral fixed points' (Cuneo & Shafer-Landau 2014: 399). These moral fixed points are thought to be conceptual truths because they are true in virtue of their constituent concepts (Cuneo & Shafer-Landau 2014: 403). Consider the moral proposition 'it is wrong to kill a child merely for fun', which is one of the obvious candidates for a moral fixed point. According to the moral fixed points theory, this moral proposition 'it is wrong to kill a child merely for fun' is true in virtue of the meanings of its constituent concepts. As mentioned above, the descriptive concept of <killing a child merely for fun> is part of the moral concept of <wrong>. Necessarily, anything that falls under the concept of <killing a child merely for fun> also falls under the moral concept of <wrong> (but not *vice versa*). Hence, the moral proposition 'it is wrong to kill a child merely for fun' is

true in virtue of its constituent concepts, and thus it is also a conceptual truth.

5.2.2 Knowing Conceptual Moral Truths

If moral truths are conceptual truths, how would we then be able to know them? Conceptual truths are often regarded as "the objects of *a priori* knowledge" (Boghossian 1997: 334), and it is an old, difficult question of how we can acquire this type of *a priori* knowledge. In this 5.2.2, I will first explain (i) how we could be argued to know conceptual moral truths and then (ii) why we could be thought to have evolved to have the competency to know the conceptual moral truths.

(i) How we can know the conceptual moral truths

Conceptual moral truths are the moral propositions that are true in virtue of their constituent concepts. Therefore, in order to know the conceptual moral truths, we need to know what the moral concepts in the moral propositions mean. And, to know the relevant meaning of the moral concepts, the first requirement is that we need to be able to use and apply those moral concepts with competence. As explained above, a speaker can apply a moral concept with competency if the descriptions that constitute the relevant descriptive definition function as the implicit principles that guide the speakers' application of the moral concept. Hence, the competency with moral concepts can also be considered to be an ability to follow the implicit principles of concept application, which can also help us to learn what the meanings of the moral concepts are (Finlay 2014: 7).

How can we, as speakers, come to learn explicitly what the implicit principles that guide

our concepts' applications are? There are two possible answers to this question. Firstly, in many cases, we have learned the meaning-constituting descriptions of the relevant concepts through being given the explicit stipulative definitions of those concepts. Many scientific, theoretical and technical concepts are introduced to our language by stipulative definitions. Let us consider the concept <carbon dioxide> as an example. We learn the description of this concept through being given the explicit stipulative definition for it. For instance, the stipulative definition of <carbon dioxide> is something like: 'It consists of two carbon atoms and an oxygen atom; at room temperature, it is a colourless gas'. In this type of cases, it is easy for the speaker to find out which principles guide her concept application and thus what the correct definition of the concept in question is. All the speaker has to do is either to remember the stipulative definition of the concept she was given earlier when she acquired the concept or by looking up the definition from a dictionary.

Yet, this method is only applicable to certain kinds of theoretical, technical and often scientific concepts. It is very likely that moral concepts are not that kind of concepts. One indication of the difference between scientific and the moral concepts is that there are rarely genuine disagreements between speakers about the meanings of most scientific concepts, which indicates that scientific concepts have often been introduced by stipulative definitions. In contrast, there are deep disagreements among speakers about the meanings of almost all moral concepts and terms. Hence, this is a reason to think that moral concepts have not been introduced by stipulative definitions, and this is why it is not convincing that we could come to know the meaning of moral concepts in the same way in which we are able to know the meaning of the scientific concepts.

The second way that we can learn the implicit principles is by doing conceptual analysis. This is also the way in which the conceptual truth realists suggest that we can know the conceptual moral truths. The question then is, what is conceptual analysis? Conceptual analysis is the method for how to make explicit the implicit descriptions that guide the way in which we apply concepts, which is done for the sake of knowing what these concepts mean. Conceptual truth realists further argue that conceptual analysis is applicable to moral concepts too. The ability of conceptually analyzing moral concepts is just an extension of our more general ability to analyze concepts.[112]

One traditional way to understand conceptual analysis is to think that it proceeds by testing descriptions that are suggested to constitute the definitions of the relevant concepts with potential counter-examples. Consider the famous example of the definition of the concept <bachelor> in terms of the descriptions '… is unmarried' and '… is male'. During the first stage of conceptual analysis, we rely on our intuitions to test whether the things that we would intuitively categorize under the concept satisfy the proposed definition. At this point, we need to imagine some potential counter-examples to see whether the relevant descriptions really can provide the proposed definition of the concept in question. For instance, the Pope satisfies both the descriptions '… is unmarried' and '… is male' even if it would be odd and counter-

[112] Cuneo and Shafer-Landau (2014) suggest that our ability to "classify some moral propositions as a species of conceptual truth" is a manifestation of the generic capacity of being able to classify other propositions as conceptual truths (427). As for the analytic naturalists, we seem to be able to conceptually analyze simple terms such as <vixen> and <bachelor> successfully. Analytic naturalists then attempt to use this method of conceptual analyses to give a reductive definition of moral concepts (i.e., <good>, <right>, <wrong>) (Fisher 2011: 60-61).

intuitive to say that the Pope is a bachelor.[113] Because of this, we should reject the relevant descriptions as the correct definition of the concept <bachelor> and also try to find another set of descriptions that better capture the meaning of the concept in question by matching which things we would classify under that concept (perhaps by adding the description '…is not the Pope' to the previous definition). Through this process of revising our definitions on the basis of testing them against potential counter-examples, we can hope to make explicit the implicit principles that govern our concept application and thus which constitute the descriptions that can be used to define our concepts reductively.

(ii) Why we have evolved to have the competency to know the conceptual moral truths

Conceptual truth realists argue that there is an evolutionary explanation of our ability to do conceptual analysis and that ability can also help us to know the conceptual moral truths. They may even concede that the competency to know the conceptual moral truths in question itself is not necessarily fitness-enhancing. Nevertheless, according to them, we have evolved to have the general cognitive capacities that are required for doing conceptual analysis on all concepts, including the simple concepts like <vixen> and <bachelor>. On their view, it is fitness-enhancing to have that kind of generic capacities that enable us to know conceptual truths.[114] Furthermore, as mentioned above, the

[113] Michael Tye (1991) uses this example of the Pope to argue that concepts do not have any necessary and sufficient conditions that can specify their meanings (144-145).

[114] Conceptual truth realists (especially Shafer-Landau and Cuneo) seem to take it for granted that it is fitness-enhancing for us to have the ability to conceptually analyze concepts. I also think that this ability is fitness-enhancing because the method of conceptual analysis provides a way for speakers to use and apply many concepts without arbitrary meaning. As a result, the ability to conceptually analyze concepts can be considered to be enhancing communications and co-operation between individuals, and thus it is fitness-enhancing (see also 2.3.2).

ability to conceptually analyze moral concepts could be argued to be just an extension or a manifestation of the previous type of general capacities. Because of this, the conceptual truth realists suggest that it is reasonable to think that we have also evolved to have the more specific ability to conceptually analyze moral concepts.

Conceptual truth realists like Cuneo and Shafer-Landau rely heavily on the analogy of advanced mathematics to explain why they think that we would have evolved to have the competency for analyzing moral concepts.[115] We have evolved to have many competencies for being able to acquire knowledge in other domains that are not necessarily fitness-enhancing in themselves. For example, our ability to appreciate advanced mathematics like set theory or topology is not fitness-enhancing (Cuneo & Shafer-Landau 2014: 427). This ability, however, is a "natural extension of more general powers of reasoning that [are] surely fitness-enhancing", that is, our more general competencies of doing simple mathematics and arithmetic that had survival value (Cuneo & Shafer-Landau 2014: 427).

Now, we can be quite sure that we have evolved the general competency of doing simple mathematics and arithmetic. Firstly, we are able to reliably acquire simple mathematical knowledge by using our mathematical skills. Secondly, and more importantly, these are genuinely fitness-enhancing skills. For example, the skill of doing arithmetic enables us to make correct predictions that can help us to survive in the wild. It is just a happy coincidence that the extension and manifestation of these skills of doing simple

[115] For an analytic naturalist's explanation of why we would have evolved to have the competency to acquire conceptual moral knowledge, see Finlay (2014: 16).

mathematics can help us to acquire also more advanced mathematical knowledge. Following the same line of reasoning, conceptual truth realists argue that we have evolved to have the required capacities for doing conceptual analysis in order to know the conceptual moral truths even if those abilities are themselves not related to species' survival or fitness. It is just that these required abilities are merely a manifestation or an extension of the more general cognitive capacities that are genuinely fitness-enhancing (for example, the basic capacities for thinking abstractly and using language).

5.2.3 Why the Epistemic Status of Moral beliefs is not Undermined by Their Evolutionary Origin

In this final sub-section of §5.2, I will finally explain why the conceptual truth realists argue that the epistemic status of our moral beliefs is not undermined by their evolutionary origin. The conceptual truth realists need not even to reject what I have argued for in Chapter 2. That is, they do not need to reject the view that evolutionary debunkers can wholly explain the evolutionary origin of our moral beliefs. Instead, the conceptual truth realists merely try to argue that there is a reliable way in which our moral beliefs can still be justified, and this way is not threatened by the evolutionary origin of our moral beliefs.

In order to know whether an agent's belief is justified, the belief should be evaluated in an epistemic sense that is relative to the aim of "maximizing truth and minimizing falsity" (Alston 1985: 59). This entails that there must be some sort of a connection

between a belief being justified and it being true (Cohen 1984: 279).[116] According to the conceptual truth realists, we have evolved to have the faculties that are required for reliably accessing the conceptual moral truths, that is, for doing conceptual analysis. If they are right, we have the competency to discover which of our moral beliefs are true by doing conceptual analysis. If this is also right, then our moral beliefs can be justified in a reliable way and furthermore in such a way that the epistemic status of our moral beliefs could not be undermined by their evolutionary origins.

For example, let us consider the utilitarian who believes that 'it is right to maximize happiness' and who is also looking for the justification for this belief. We can also imagine that she uses the method of conceptual analysis described above to analyze the moral concept <right>. If she then found out through the analysis that the description '… maximizes happiness' really captures the meaning of <right>, she would be justified to believe that 'it is right to maximize happiness'. After all, she will now have good reasons to think that this moral belief is very likely to be true.

The conceptual truth objection thus argues that the evolutionary explanation of our moral beliefs fails to undermine the justification of our moral beliefs. Even if our moral beliefs do have an evolutionary origin that is considered to be unreliable, those moral beliefs can be justified by relying on a reliable, truth-tracking method, that is, by relying on conceptual analysis. More importantly, the conceptual truth realists also explain how we have evolved to have the competency to conceptually analyze moral concepts, and

[116] Internalists and externalists concerning justification both agree that there must be some sort of a connection between a belief being justified and it being true. See Feldman (2003: 39-107).

they also argue that our moral beliefs can be justified in this reliable way that is not threatened by any concerns about the evolutionary origin of these moral beliefs.[117]

5.3 Reply 1: Evers and Streumer's Objection to the Moral Fixed Points Theory

The aim of this Chapter 5 is to defend the EDA against the Conceptual Truth Objection. Hence, in the rest of this Chapter, I will introduce and evaluate three potential ways that evolutionary debunkers could deal with the objection in question. Firstly, §5.3 will introduce Evers and Streumer's objection (2016) to the moral fixed points theory. However, in this §5.3, I will also argue that their objection fails to help evolutionary debunkers to reject the conceptual truth objection because it is not a challenge that is applicable to all versions of conceptual truth realism, or so I will argue. Then, in §5.4 and §5.5, I will introduce two better reasons that which evolutionary debunkers have for rejecting the Conceptual Truth Objection.

Hence, in the rest of this section §5.3, I will focus on Evers and Streumer's objection (2016) to the moral fixed points theory. We can start from the following generic formulation that captures the basic crux of their objection:

Evers and Streumer's argument against the moral fixed points theory

1. All conceptual moral truths are moral propositions that are true merely in

[117] Therefore, the conceptual truth realists may agree that the EDA actually works as an objection against other moral realist theories. Nevertheless, they are not responsible for defending those theories as they merely aim to argue that the EDA fails to pose an epistemological challenge to conceptual truth realism.

virtue of their constituent concepts (definition from the conceptual truth realism, including the moral fixed points theory).

2. For any moral propositions to be true, the constituent moral concepts of the moral propositions need to ascribe moral properties that are instantiated.

3. If the constituent moral concepts of the moral propositions ascribe moral properties that are instantiated, then those moral propositions cannot be true merely in virtue of their constituent concepts.

4. No moral propositions are true merely in virtue of their constituent concepts. (2, 3, *Modus Ponens*)

Conclusion: There are no conceptual moral truths. (1, 4, *Modus Tollens*)

As I explained in the previous §5.2, the 1st premise is thought to be true according to the conceptual truth realism, including the moral fixed points theory. Let us then consider why Evers and Streumer think that the 2nd, 3rd and 4th premises too are true. We can start from the 2nd premise and use the concept <God> as an example. Evers and Streumer (2016: 3) imagine a Christian who believes the proposition 'benevolence is rewarded by God'. Assume that this Christian also wants to show that the previous proposition is a conceptual truth and so she needs to show that the proposition is true merely in virtue of its constituent concepts. According to the moral fixed points theory, she would thus need to know the essence of the constituent concepts of the proposition

in question.

Nevertheless, Evers and Streumer argue the fact that this Christian can analyze and know the essence of the concept <God> does not necessarily mean that God exists. What she really gets from the essence of the concept <God> is only the conditional proposition 'if anything is rewarded by God, benevolence is rewarded by God' that is true mere in virtue of its constituent concepts (Evers & Streumer 2016: 4). Hence, even if the Christian knew the essence of the concept <God>, she would not thus know whether the proposition 'benevolence is rewarded by God' too is true because she may not know whether God exists.[118]

Following the same line of reasoning, we can then consider the moral proposition 'it is wrong to kill a child merely for fun' as an example again. Let's us again assume that, according to the essence of the moral concept <wrong>, anything that falls under the concept <killing a child merely for fun> also falls under the moral concept <wrong>. Nevertheless, even if we knew the essence of the moral concept <wrong>, this would not mean that the moral proposition in question could also be true merely in virtue of its constituent concepts. Instead, as a consequence of the essence of the moral concept <wrong>, merely the conditional moral proposition 'if anything is wrong, it is wrong to kill a child merely for fun' would be conceptually true.

Moreover, even if that conditional moral proposition were true, this would not thus

[118] In *Proslogion* (1059/1998: 87-88), Anselm of Canterbury introduced an argument to prove the existence of God from the concept of <God>. This argument is known as the 'ontological argument'. For overviews of the ontological argument, see Matthews (2005: 81-102) and Rowe (1978/2017: 36-46).

entail that it is really wrong to kill a child merely for fun. The moral proposition 'it is wrong to kill a child merely for fun' can be true only if there exists of an instantiated moral property of wrongness in the reality that can be ascribed by the constituent moral concept <wrong> in the previous moral proposition. This too means that, even if we knew the essences of the moral concepts that constitute moral propositions, we would still not know which moral propositions (if any) were true. Hence, according to Evers and Streumer, a moral proposition can be true only if the constituent moral concept of that moral proposition ascribes the moral property that is instantiated in reality. The 2nd premise is thus argued to be true.

We can then consider the 3rd premise. According to the previous 2nd premise, whether a moral proposition is true depends on whether the constituent moral concept of that moral proposition ascribes a moral property that is really instantiated. Evers and Streumer (2016) then argue that, if the constituent moral concepts of moral propositions ascribe moral properties that are instantiated, it can be argued that moral propositions are actually made true by those moral properties in the reality (7). As a result, those moral propositions cannot be true merely in virtue of their constituent concepts, and thus the 3rd premise should also be considered to be true. If the 2nd and 3rd premises are both true, then we can infer that no moral propositions can be true merely in virtue of their constituent concepts by *modus ponens*. The 4th premise is therefore thought to be true. Hence, if the 1st and 4th premises are both true, then we can infer that there are no conceptual moral truths by *modus tollens*, or so Evers and Streumer argue.

So far so good. I agree that Evers and Streumer's argument can be considered to be a

genuine objection to the moral fixed points theory. Nevertheless, I believe that, despite this, their argument is not a decisive objection to the Conceptual Truth Objection because it fails to pose a challenge to all versions of conceptual truth realism.[119] It is especially unclear how their objection could pose a genuine challenge against the analytic naturalism. In the rest of this §5.3, I will then argue that the analytic naturalists can be thought to be rejecting the 3rd premise of the previous argument, that is, the conditional 'if constituent moral concepts of the moral propositions ascribe moral properties that are instantiated, then the moral propositions cannot be true merely in virtue of their constituent concepts'.

To explain why the analytic naturalists can plausibly reject the previous entailment, we can start from why the moral fixed points theorists need to consider that entailment is true. Let us assume that the 2nd premise of the Evers and Streumer's argument is true and therefore moral propositions can be true only if their constituent moral concepts can ascribe moral properties that are really instantiated. The defenders of the moral fixed points theory, including Shafer-Landau (2003), believe that these ascribed moral properties are additional, *sui generis*, non-natural moral properties (65-67). In this situation, the defenders of the moral fixed points theory will find it difficult to show that there exist non-naturalistic moral properties in the world to which the constituent moral concept refers.[120]

[119] In addition, David Killoren (2016) also suggests that Evers and Streumer's objection is not a decisive reason to reject the moral fixed points theory. As he claims, "it is possible [for the defenders of the moral fixed points theory] to imagine [a] response" to their objection (Killoren 2016: 173). Unfortunately, Killoren does not further explain what the response could be.

[120] Evers and Streumer (2016) also mention that the moral fixed points theory is less attractive than robust moral realism because of its lack of ontological commitment (6n13). But they do not actually explain why the moral fixed points theory is thus less attractive.

Let us consider the moral proposition ‘it is wrong to kill a child merely for fun’ as an example again. Once again, let’s assume that necessarily, anything that falls under the concept of <killing a child merely for fun> also falls under the moral concept of <wrong>. However, even speakers who can apply the moral concept <wrong> with competency cannot really know whether the moral proposition in question is true if we assume that non-naturalist realism is true. This is because they cannot know on the basis of their conceptual competency whether the relevant non-natural moral property of wrongness really exists. Thus, the defenders of the moral fixed points theory are unable to show why the meaning of the constituent moral concept in the previous moral proposition would ensure that that proposition is true.

In contrast, even if the analytic naturalists thought that the 2nd premise of the Evers and Streumer’s argument were true, they can still reject the 3rd premise of that argument. This is because, on their view, moral propositions can still be true merely in virtue of their constituent concepts even if these constituent moral concepts ascribed moral properties that are a part of the reality. To recall, analytic naturalists claim that moral properties can be reduced to naturalistic properties, which are such that no one doubts their existence in this debate. Hence, they can also argue that moral propositions can be true given that there is a realm of natural moral properties in the world that the constituent moral concepts of those moral propositions really refer to. If they are right, then, according to the general definition of conceptual truth realism, it is a part of the meaning of the moral concepts that some actions do have relevant moral properties that really exist in reality. If we, on this view, knew what the constituent moral concepts in

the moral propositions mean, we would also know whether those moral propositions are true.

Let us consider the already mentioned moral proposition ‘it is right to maximize happiness’ and assume, only for the sake of the argument, that the moral concept <right> means ‘maximizes happiness’. According to analytic naturalism, if speakers knew the meaning of the moral concept <right> which refers to the natural property of ‘maximizing happiness’, they would also know that the moral proposition in question is true – there would be no meaningful questions about the existence of the property of rightness which the predicate ‘is right’ ascribes to the actions in this framework. As a result, analytic naturalists can plausibly reject Evers and Streumer’s objection by denying the 3rd premise of their argument. This is because, on their view, even if the constituent moral concepts of the moral propositions ascribe moral properties that need to be instantiated for the propositions to be true, those moral propositions could still be merely true in virtue of the concepts used.

In fact, Evers and Streumer (2016) themselves also admit that moral propositions could still be true merely in virtue of the constituent concepts if these concepts ascribed natural moral properties that are non-problematically instantiated in the reality (6). In response, they might insist that their objection is a targeted argument, which merely aims at showing that the moral fixed points theory is false in the non-naturalist framework. As long as their objection can pose a real threat to the moral fixed points theory, the argument has already achieved its aim. That’s true. The aim of this Chapter 5, however, is to investigate the ways in which the evolutionary debunkers could reject

the Conceptual Truth Objection. Unfortunately, Evers and Streumer's objection fails to help the evolutionary debunkers reject an important version of conceptual truth realism, that is, the analytic naturalism. As a result, their objection can at best work as a partial response from the debunkers to the conceptual truth realism.

In §5.4 and §5.5, I will instead provide two replies to the Conceptual Truth Objection that are both meant to be objections to all versions of conceptual truth realism. Unlike Evers and Streumer's argument, my replies below can help evolutionary debunkers reject the Conceptual Truth Objection in a more convincing and general way.

5.4 Reply 2: A Hare-style Argument Against the Application of the Classical Theory of Concept in Morality

In this §5.4, I will introduce an R.M. Hare-style argument (1952/1991) against the idea that the classical theory of concepts is applicable to moral vocabulary. As I mentioned in §5.2, conceptual truth realism heavily relies on that theory to explain how moral truths are supposed to be conceptual truths. Hence, if the Hare-style argument is sound, the classical theory of moral concepts must be considered to be false, and the Conceptual Truth Objection must also be rejected as a result. This §5.4 has three parts. 5.4.1 will first explain the basic crux of the Hare-style argument and why it is a sound argument. In that sub-section, I will conclude that the Conceptual Truth Objection should therefore be rejected because of that argument. Then, in 5.4.2 and 5.4.3, I will introduce and also reject two possible objections to the Hare-style argument, which are suggested by Frank Jackson (2008) and Philippa Foot (1958) respectively.

5.4.1 The Basic Crux of the Argument

The R.M. Hare-style argument (1952/1991) against the application of the classical theory of concepts to moral concepts can be thought to be an application of his imaginary case of the missionary and the cannibals (148-150). By introducing this case, Hare aims to argue that the classical theory of concepts must be false when it comes to moral concepts. If Hare is right, it would be implausible for the conceptual truth realists, who attempt to apply the classical theory of concept to moral concepts, to argue that moral truths are conceptual truths. The basic crux of the Hare-style argument can be formulated as follows:

> **A Hare-style Argument against the Application of the Classical Theory of Concepts to Moral Concepts**
>
> 1. Whichever implicit descriptive principles guide a speaker's use of a moral expression determines which moral concept the speaker is applying (the classical theory of concepts).
>
> 2. Different implicit descriptive principles can guide different speakers' application of the same moral expression.
>
> 3. Two speakers who are guided by different implicit descriptive principles of application of the same moral expression will not possess the same moral concept. (1, 2)

4. If two speakers use the same moral expression to express different moral concepts, they cannot use this moral expression to agree or disagree.

5. Two speakers can use the same moral expression to agree and disagree even if different implicit descriptive principles governed their use of the moral expression.

6. Two speakers, who use the same moral expression but are guided by different implicit descriptive principles, do not possess different moral concepts. (4, 5, *Modus Tollens*)

7. The 3rd Premise is false. (6, Contraction)

8. The 1st Premise is false. (3, 7)

Conclusion: The classical theory of concepts must be false when it comes to moral concepts. (from the denial of 1st premise)

Let me then explain the premises of the argument above starting from the 1st premise. As explained in §5.2, according to the classical theory of concepts, when a competent speaker uses a moral expression, she must be guided by an implicit principle that corresponds to a descriptive definition. Moreover, this implicit principle also determines which moral concept she is applying when using the moral expression in question. Hence, according to the conceptual truth realists and the defenders of the

classical theory of concepts, which are the targets of the Hare-style argument, the 1st premise is true.

Then, in order to explain why the remaining premises too are true, I substantially draw on R.M. Hare's (1952/1991) famous example of the missionary and the cannibals. Let us begin from the 2nd premise. In *The Language of Morals*, Hare imagines that a missionary arrives at a cannibal island with a grammar book. The missionary knows a word in the cannibals' language, which is equivalent to the English word 'good', according to his grammar book (Hare 1952/1991: 148). Let's assume that the grammar book is correct and that both the missionary and cannibals are using the English word 'good' and the equivalent word in the cannibals' language as "the most general adjective of commendation in their language" respectively (Hare 1952/1991: 148). Hence, when the missionary uses the word 'good', the cannibals are able to know that he is using the word to commend a person or an object, and *vice versa* (Hare 1952/1991: 148).

Nevertheless, in Hare's example (1952/1991), the missionary applies that English word 'good' to people who are "meek and gentle", while cannibals apply to equivalent word in their language to people who "collect more scalps than the average" (148). In this situation, the missionary and the cannibals thus have different descriptive principles in mind when they use and apply the same moral word 'good'. Therefore, different implicit principles and descriptions can guide the speakers' application of the same moral term – 'good'. As a result, the 2nd premise can be argued to be true.

If the 1st and 2nd premises of the argument in question are true, then the 3rd premise too

should be thought to be true. To illustrate this, we can start from considering the word ‘bank’ as an example. Frank asks Emily to meet him near a bank. While Frank thinks that they are going to withdraw some money from the ‘bank’ in the city centre, Emily thinks that they will go fishing at the river ‘bank’. In this case, Frank and Emily are guided by different implicit principles when they apply the term ‘bank’ to different objects. Because the term ‘bank’ is ambiguous in this way, we can think that it expresses different concepts depending on what description the speaker has in mind. This means that, in this case, Frank and Emily are not relying on the same concept when using the same word ‘bank’.

Following the same line of reasoning, the defenders of the classical theory of concepts too would need to concede at this point that the missionary and the cannibals do not possess the same moral concept. This is because they are guided by different principles when they use and apply the same word ‘good’, just like in the previous example of the term ‘bank’. While the cannibals would use the term ‘good’ to commend people who collect most scalps by killing others, the missionary would use the same term ‘good’ to commend people who do not murder others for scalps. Although they seem to use the same term, they would not possess the same moral concept <good> because they are guided by different principles that determine the way that they use that term. Thus, the 3rd premise can be thought to be true, provided that the 1st and 2nd premises of the Hare-style argument too are true.

Let me then explain the motivation behind the 4th premise. As we saw above, according to the classical theory of concepts, Frank and Emily do not have the same concept in

mind when they both use the term ‘bank’ in the previous example. Let’s further assume that Frank has in mind the concept <bank$_1$>, which refers to a financial establishment whereas Emily has in mind the concept <bank$_2$>, which means a land alongside a river. In this situation, they are just talking past each other when it comes to the term ‘bank’. This also means that any communication between them regarding that term will fail. Frank cannot understand why Emily wants to go fishing at the bank, while Emily also cannot understand why Frank wants to withdraw some money from the bank.[121]

Following the same line of reasoning, if two speakers use the same moral expression to express different moral concepts, they will be simply talking past each other. Any communication or moral disagreement between them based on their attempts to use the moral expression in question would simply fail. One would be talking about oranges and one about apples. As a result, the 4th premise can be argued to be true, since both moral agreement and disagreement between two speakers is possible only if they use the same expression to express the same moral concept.

We can then consider the 5th premise. According to Hare (1952/1991), in the imaginary case of the missionary and the cannibals, they are intuitively not talking past each other (148). When the missionary applies the moral concept <good> to certain people and

[121] Let us imagine another situation where Frank and Emily now have a phone conversation. We can also imagine that they now notice that they do not have the same thing in mind even if they are both using the same term. Let’s imagine that Frank realizes that Emily has applied the concept ‘bank’ to land alongside a river rather than to a financial establishment. In this case, he will then be guided by the implicit principle that also guides Emily’s application of the term ‘bank’. In other words, now he possesses both concepts <bank$_1$> and <bank$_2$>. Furthermore, he also knows why Emily thinks that they are going to meet at next to a river rather than at the financial establishment. As a result, communication is possible in this example only if Frank and Emily are able to translate each other’s utterances to their own language. After the translation, however, they will both be able to use the expression ‘bank’ to express the same concept <bank$_2$> in their further discussion.

actions, the cannibals understand that he wants to commend some people, and *vice versa*. Even if the missionary and the cannibals were relying on different application conditions when using the moral term 'good' and were applying the term to different people, they both use the same moral concept <good> to commend others. This is guaranteed by the same practical role that the moral concept <good> plays in their practical reasoning, that is, to commend people and objects. For instance, they would both think that people should be motivated to do good actions and people who don't do good actions should be criticized. Hence, both the missionary and cannibals still possess the same moral concept <good> even if they are guided by different descriptive principles. As a consequence, two speakers who are able to use the same the moral expression should also be able to use this expression to communicate: to both agree and disagree even when they are guided by different implicit descriptive principles of the application of that moral concept. Thus, the 5th premise should also be accepted to be true.

Following the 4th and 5th premises, the 6th premise too is true according to the inference rule *Modus Tollens*. But, if the 6th premise is true, then it will contradict the 3rd premise of the argument in question. The 3rd premise is thought to be true according to the defenders of the classical theory of concepts, who are committed to thinking that the missionary and the cannibals possess different moral concepts because they are guided by different implicit principles of the application of the moral concept <good>. However, the missionary and the cannibals are able to use the moral term 'good' to communicate about morals as we just saw and, as a consequence, they cannot possess different moral concepts. This is because they are able to psychologically grasp the

evaluative meaning of the moral concept <good> – that the concept is "the most general adjective of commendation" – when they use the moral concept in question (Hare 1952/1991: 149). As a result, two speakers using the same moral expression who are guided by different implicit descriptive principles are able to possess the same moral concept, and thus the 3rd premise is likely to be false.

However, the 3rd premise of the Hare-style argument could only be true if the 1st and 2nd premises of that argument were both true. This means that, if the 3rd premise is false, then at least one of the 1st and 2nd premises must be false. Moreover, it is very implausible to reject the view that speakers can be guided by different descriptive principles of application of a moral expression. Hence, the 2nd premise seems to be true, and therefore the 1st premise, which is the core definition of the classical theory of concepts, must be false. If the 1st premise is false, this also implies that the classical theory of concepts must be considered to be wrong at least when it is applied to moral concepts. The classical theory of concepts may, of course, be true in other domains, but at least if the Hare-style argument above is sound then that theory must be false when it comes to moral concepts. As a result, the conceptual truth realists will not be able to claim that moral truths are conceptual truths and furthermore the Conceptual Truth Objection must be rejected because the classical theory of concepts on which it is based on turns out to be false.

5.4.2 Response to Jackson's Objection

In the remaining part of §5.4, I will consider two objections to the Hare-style argument against the application of the classical theory of concepts to moral terms. Firstly, in this

sub-section 5.4.2, I will focus on the first objection suggested by Frank Jackson in his article "The Argument from the Persistence of Moral Disagreement" (2008). In that article, Jackson, who also is an analytic naturalist, explicitly argues that, even if two speakers used the same words to express different moral concepts, they would still be able to communicate and to disagree about moral question. As a result, his objection can be considered to be a challenge against the 4th premise of the Hare-style argument.

For the sake of defending his own view, Jackson (2008) borrows the expressivists' account of moral disagreement to argue that there is a way in which two speakers who employ different moral concepts can disagree (75-77). Let us then start from considering what the expressivists' account of moral disagreement is. Let's imagine an example in which Harry and Mary are debating whether we should eat meat. In this example, Harry thinks that it is morally wrong to eat meat, whereas Mary thinks that there is nothing wrong about meat eating. Let us also assume that they are both well informed, and there is no disagreement between them on any non-moral facts related to this issue.[122]

In this example, expressivists argue that there is a genuine disagreement between Harry and Mary. On their view, moral claims and judgments are expressing pro- and con-attitudes rather than beliefs that represent moral facts or properties in the reality. When Harry claims that 'it is morally wrong to eat meat', he simply means something like 'Boo! Meat eating!' – he is expressing his disapproving attitudes. Likewise, when Mary

[122] For example, we can assume that they both know the relevant empirical facts about how animals suffer in the meat industry.

suggests that ‘it is not morally wrong to eat meat’, she means something like ‘Hurray! Meat eating!’ – she is expressing her approving attitudes. Thus, expressivists understand moral disagreement as disagreement in the expressed attitudes. According to their view, this means that Harry and Mary have a genuine moral disagreement because their attitudes towards meat eating conflict.[123]

We may further ask what a conflict in attitudes is. Let us consider the following example of deciding where to have dinner. Suppose that Aki and Julia are now discussing where to have dinner tonight as that they really want to have dinner together. However, Aki wants to go to a Chinese restaurant, whereas Julia wants to go to an Indian restaurant instead. Let’s also assume that they have no disagreement about any related facts, such as how busy the restaurants are. In this situation, they have conflicting attitudes towards where they are to have our dinner tonight. Yet, intuitively, there is a genuine disagreement between Aki and Julia because they have conflicting preferences or plans of where they are to eat. If they really want to have dinner together, at least one of them needs to give up her plan or preference. In other words, at least one of them is required to change the attitudes of the other or simply change her own attitude (Wong 1984: 10). This also means that, in this situation, they are not talking past each other. This is because their disagreement in their plans of where to have dinner is not merely a verbal or an apparent one.

We can then return to the previous of example Harry and Mary. According to expressivism, Harry and Mary are also not talking past each other because their moral

[123] For a representative objection to this view of disagreements, see Stevenson (1945: 1-19).

claims expressing their conflicting attitudes towards meat eating. Hence, when they argue about whether 'eating meat is morally wrong', they have conflicting attitudes exactly in the same way in which Aki and Julia have a conflict in their plans or preferences concerning where to have dinner. After all, the disagreement between Harry and Mary is not a factual one. Instead, their disagreement reflects a deeper conflict in their attitudes towards meat eating. For this reason, expressivists think that there is a genuine moral disagreement between Harry and Mary.

Let us then consider how Frank Jackson borrows the expressivists' account of moral disagreement in order to defend analytic naturalism. He starts by suggesting that this account of moral disagreement is not exclusive to expressivism. When it comes to moral disagreements, he argues that subjectivists too can understand such disagreements as disagreements in attitude rather than as factual disagreements (Jackson 2008: 84).[124] Unlike the expressivists who think that moral utterances are merely expressing practical attitudes, subjectivists, as cognitivists, argue that those utterances are instead reports of our attitudes (in fact, reports of the very same attitudes which the expressivists take to be expressed by those utterances) (Miller 2003: 37; Köhler 2012: 73). This also means that subjectivists can further argue that two individuals can use a moral term to disagree even if they use that term to report their attitudes in a way that does not constitute a genuine factual disagreement as long as

[124] In an earlier article "A Problem for Expressivism", Frank Jackson and Philip Pettit (1998) suggest that almost all meta-ethical theories "can respond to the problem of moral disagreement simply by noting that a difference in moral attitudes can survive agreement over all the facts" (251). However, in a more recent article "The Argument from the Persistence of Moral Disagreement" (2008), Jackson suggests that subjectivists are more likely to adopt the expressivists' account of moral disagreement (77). Likewise, Björnsson and Finlay (2010) argue that meta-ethical contextualists, who accept a kind of relativism in ethics, can adopt the expressivists' account of moral disagreement (27-28).

they are reporting conflicting attitudes towards the same moral issue (Jackson 1998: 162; 2008: 77).

Let's return to the example of Harry and Mary again. For the subjectivists, when Harry suggests that (1) 'it is morally wrong to eat meat', he means that 'I disapprove of meat eating'. Likewise, when Mary claims that (2) 'it is not morally wrong to eat meat', she means 'I approve of meat eating'. The thing to note is that both of these claims can be true at the same time and so there is no factual disagreement between Harry and Mary. Yet, the subjectivists can argue that Harry and Mary are really disagreeing nevertheless: They are disagreeing in the way that expressivists suggest they are really disagreeing. This is because, according to the subjectivists, the moral utterances (1) and (2) are really reporting the conflicting practical attitudes of Harry and Mary towards meat eating respectively. Hence, they can argue that Harry and Mary have a genuine disagreement about meat eating that, instead of a factual disagreement, consists of a disagreement in attitude towards meat eating (exactly in the same way in which expressivists understand their disagreement). As a consequence, according to Jackson, subjectivists too can adopt expressivists' account of moral disagreement.

We can now consider why Jackson thinks that the 4th premise of the Hare-style argument is false. Let us return to the example of the missionary and cannibals again. Let's suppose that the missionary now claims that 'kindness is good', whereas the cannibals suggest that 'kindness is not good'. According to the subjectivists, the missionary actually means that 'I approve of kind actions or persons', whereas the cannibals mean that 'we disapprove of kind actions or persons'. If the subjectivists are

right in how those utterances should be understood, the previous claims can of course both be true at the same time because they just report their different attitudes toward kindness. Thus, even if in this case there is no factual disagreement, the subjectivists can count the difference in the attitudes toward 'kindness' as a disagreement (Jackson 2008: 82). Hence, even if two speakers use the same term 'good' to ascribe different properties, they can still, according to Jackson, use the moral term in question both to communicate and also to disagree as long as there is a relevant disagreement in attitude between them. As a result, according to Jackson, the 4th premise of the Hare-style argument should be rejected.

Nevertheless, I want to now argue that Jackson's objection is not a decisive challenge to evolutionary debunkers who adopt the Hare-style argument with the aim of rejecting conceptual truth realism. This is because Jackson's objection is a targeted argument that attempts to defend both analytic naturalism and subjectivism (Jackson 2008: 76-77). Evolutionary debunkers therefore can argue that the conceptual truth realists face a dilemma: Either (1) they adopt Jackson's objection but the cost of this is having to give up their realism and adopt subjectivism instead, or (2) they do not give up realism and so they need to look for some other response. As a consequence of this dilemma, I argue that Jackson's objection fails to help the conceptual truth realists to give a plausible response to evolutionary debunkers who adopt the Hare-style argument.

Let me start with explaining the first horn of the dilemma. As we saw above, the conceptual truth realists can adopt Jackson's response only if they can understand moral disagreement as disagreements in attitude rather than as factual disagreements. For the

sake of simplicity, I will call those conceptual truth realists who understand moral disagreement as disagreements in attitude 'DIA conceptual truth realists'. We can then consider the example of the missionary and cannibals again. Let's too suppose that the missionary claims that 'kindness is good' and the cannibals suggest that 'kindness is not good'. Since DIA conceptual truth realists aim at counting the difference in attitudes toward 'kindness' between the missionary and cannibals as a moral disagreement, they thus need to analyze the moral concept <good> in terms of the speaker's attitude towards 'kindness'. More precisely, on their view, the moral concept <good> is thus defined with the description 'the speaker has a positive attitude toward certain things or acts'. As a result, the DIA conceptual truth realists can argue that the missionary and cannibals are able to use the same moral concept <good> to disagree as long as they are using that concept to report conflicting attitudes towards the same moral issue (in this case, kindness).

As this example illustrates, however, the resulting form of DIA conceptual truth realism would clearly entail moral subjectivism, and so it would not be a form of realism in the first place. We can focus on the missionary's moral judgment that 'kindness is good'. As just mentioned, the DIA conceptual truth realists would suggest that the moral concept <good> is analyzed in terms of the attitude of speaker who applies that concept to certain things or acts. Hence, their analysis of what the missionary means by his moral judgment that 'kindness is good' is that he has a positive attitude towards kindness. Nevertheless, this analysis can be considered to be basically the same as the subjectivists' analysis of what the missionary means by his moral judgment. According to the subjectivists, the moral judgment in question reports that the missionary's

positive attitude towards kindness. Therefore, in order to adopt Jackson's objection to the Hare-style argument, the conceptual truth realists are required to understand moral disagreement as disagreements in attitude, which requires giving up their realism and accepting subjectivism instead. This is because adopting subjectivism in the analyses of the moral terms is the only way to guarantee that the speakers in the relevant cases have conflicting attitudes towards the things that they are talking about.[125]

Let us then move onto the second horn of the dilemma. We can start from considering the conceptual truth realists who want to remain realists. The problem is that, as moral realists, they actually cannot understand moral disagreements as disagreements in attitude as there is no way for them to guarantee that speakers have conflicting attitudes in the relevant cases. To see this, let's consider the case of the cannibals and missionary again. To recap, in this case, the missionary claims that 'kindness is good', whereas the cannibals suggest that 'kindness is not good'. According to the conceptual truth realists, the missionary and the cannibals only use the same moral term to express different moral concepts as they have different application principles based on their own cultures. Hence, there is no factual disagreement between the missionary and the cannibals. According to Jackson, there is still a moral disagreement between them as long as there is a relevant disagreement in attitude between them.

[125] Moreover, if the DIA conceptual truth realists give up realism and accept subjectivism, they should not be considered to be one of the targets of the EDA. As mentioned in 3.2.1, the EDA is a targeted argument. For example, in her article "A Darwinian Dilemma for Realist Theories of Value", Sharon Street (2006) suggests that the target of her version of the EDA is the "realist theories of value", which are all committed to the idea that there are attitudes-independent evaluative truths (110). Hence, at least on Street's view, the DIA conceptual truth 'realists' – who adopt Jackson's response and thus understand moral disagreement as disagreements in attitudes – should not be considered to be one of the targets of the EDA in the first place. This is because the EDA is clearly not an argument that is intended to attack subjectivism – after all, Street (2016) herself also uses the EDA to motivate a form of subjectivist constructivism.

The problem is that the conceptual truth realists who want to remain as realists actually cannot guarantee that there must be a disagreement in attitude in the previous case in the way described above. This is because, on the moral realists' view, whether a thing is morally good or bad does not depend on the attitude of any individual speaker. Indeed, in the case under investigation, the missionary doesn't need to have a positive attitude towards kindness when he says that 'kindness is good' – rather all he has to be thinking is that acts of kindness satisfy certain descriptive criteria encoded in their application rules of the term. Likewise, the cannibals do not need to have a negative attitude towards kindness when they claim that 'kindness is not good', as all they have to be thinking is that acts of kindness fail to satisfy some other descriptive criteria determined by their application rules of the term. This is why, in this case, on this realist analysis there need not be a conflict in attitudes between the speakers. As a result, unless the conceptual truth realists accept subjectivism and so give up their realist commitments, they fail to guarantee that the missionary and the cannibals must have conflicting attitudes when they use the term 'good' to talk about kindness.

In conclusion, there is thus a dilemma for the conceptual truth realists who attempt to deal with the evolutionary debunkers' Hare-style argument by using Frank Jackson's response. If they take the first horn, they can adopt Jackson's response to Hare-style argument, which requires understanding moral disagreements as disagreements in attitude. The consequence of this horn, however, is that the conceptual truth realists would need to accept subjectivism and give up their realism. If, in contrast, the conceptual truth realists take the second horn which consists of remaining realists but

at the cost of not being able to rely on Jackson's response, then they will fail to respond to the Hare-style argument against the conceptual truth realism as they have no way of making sense of the intuitive moral disagreements. As a consequence of the dilemma, Jackson's objection cannot help the conceptual truth realists to make a plausible objection to the Hare-style argument.

5.4.3 Response to Foot's Objection

We can now consider another possible objection to the Hare-style argument, which is suggested by Philippa Foot (1958) in her article "Moral Beliefs". Her objection can be considered to be an objection to Hare's view of moral concepts. According to Foot (1958), if Hare's view on moral concepts were right, then competent speakers could apply moral concepts to everything whatsoever, including very odd and bizarre things (84-85). In order to avoid this unwanted outcome, we should not accept Hare's view on moral concepts and as a consequence reject the Hare-style argument as well, or so Foot argued.

Let us start from considering Hare's example of the missionary and the cannibals again. According to Hare, they possess the same moral concept <good>, provided that they are using that moral concept in the same evaluative sense, that is, to commend others. Nevertheless, Foot argues that this view of the moral concept <good> would eventually allow competent speakers to apply the concept under investigation to anything whatsoever. This is because, on Hare's view, a competent speaker possesses a moral concept as long as she can use that moral concept in a way that fulfils its practical and action-guiding role (Foot 1958: 88). For instance, it would be very odd to think that a

competent speaker can call another person good simply because she clasps and unclasps her hands (Foot 1958: 84-85). However, if Hare's view on moral concepts were right, then a speaker who said such things is still thought to be possessing the moral concept <good> even if she applied this moral concept to the person who performs a "trivial and pointless action" such as the one mentioned (Foot 1958: 92).

Hence, Foot (1958) argues that Hare's view of moral concepts is mistaken because there are not enough restrictions on the use and application of moral concepts on that view (85). It is very reasonable to think that speaker who applies the moral concept <good> to odd actions like clasping hands should not be considered to be a competent speaker who knows what the moral concept <good> really means. On Foot's view of moral concepts, speakers can possess a moral concept only if they have at least some reasonable substantial descriptions in mind when they use the moral concept in question. If Foot is right, then a competent speaker who possesses the moral concept <good>, for example, could not apply it to the actions like clasping one's hand and collecting many scalps. This is because odd actions of this kind do not fit any plausible descriptive meanings of the moral concept <good>, including the likes of fulfilling a duty, performing an act of charity or doing anything that is related "virtues which we recognize" (Foot 1958: 93).

Let us then consider how Foot's view on moral concepts can be argued to be an objection to the Hare-style argument. As I mentioned in the sub-section 5.4.1, the 3rd premise of the Hare-style argument would be true if the 1st and 2nd premises of the argument in question were true. Furthermore, according to the 7th premise of the

argument under investigation, the 3rd premise is likely to be false. If the 3rd is premise false, at least one of the 1st and 2nd premises must be false. According to the Hare-style argument, the 1st premise is more likely to be false than the 2nd premise. Now, Foot would probably accept that the 3rd premise of that argument is false as well, but she would probably claim that the 2nd premise of the same argument is false. That is, she would rather reject Hare's view that speakers can be guided by different descriptive principles of application of the same moral expression, instead of rejecting the 1st premise of the argument in question (i.e., the crux of the classical theory of concepts).

As just mentioned above, Foot argues that speakers can possess a moral concept only if they have at least some reasonable substantial descriptions in mind when they use the moral concept in question. This too means that competent speakers who possess the same moral concept cannot be guided by very different substantial principles of application of the same moral expression. Thus, according to Foot, the 2nd premise of the Hare-style argument could be argued to be false. This is because, if that premise were true and speakers could really be guided by very different descriptive principles of application of a moral expression, then it seems that there would not be enough restrictions on the use of this expression. As a result, we would also need to accept the unwanted consequence that competent speakers could apply the moral concept to anything whatsoever, including very odd and bizarre things. If the 2nd premise of the Hare-style argument really is false, as Foot has argued, then the 1st premise of that argument and also the classical theory of concepts, as applied to moral concepts, could still be true. This would mean that the Conceptual Truth Objection that relies on the classical theory of concepts could still work as an argument against the EDA.

In defence of the Hare-style argument and the EDA, I argue that the evolutionary debunkers need not disagree with Foot's view on moral concepts. However, I argue that the 2nd premise of the Hare-style argument (i.e., different implicit principles can guide different speaker's application of a moral expression) can still be argued to be true even if we accept Foot's view of moral concepts. This means that the Hare-style argument can still work as an objection to the Conceptual Truth Objection even if we accept the basic spirit of Foot's view of moral concepts. In other words, Foot's response to Hare does not really provide what the conceptual truth realists would need for their purposes.

Why does Foot's view of moral concepts not entail that the 2nd premise of the Hare-style argument is false? Firstly, the evolutionary debunkers need not disagree with Foot that competent speakers must be thought to have at least certain reasonable substantial descriptions in mind when they use the moral concepts. They can also agree that this kind of substantial descriptions will be sufficient to guarantee that competent speakers will not apply the moral concepts to odd and bizarre things. Nevertheless, even if we should be thought to have certain substantial descriptions in mind when we use, for example, the moral concept <good>, this still leaves us with a wide range of alternatives concerning what things can be good.

Let us consider why Foot thinks that the moral concept <good> should not be applied to the act 'clasping one's hand' again. This is because she thinks that odd actions like that do not fit any plausible descriptive meanings of the moral concept <good>, including the likes of fulfilling a duty or performing an act of charity (Foot 1958: 93).

However, this actually means that, on Foot's view, there is still a wide range of plausible descriptive principles that can guide speakers' application of the moral expression <good>. Different competent speaker can still be guided by many different plausible descriptive meanings of the moral concept <good> just as long as those descriptions satisfy Foot's constraints.

We can then consider the consequence of there really being many plausible descriptive principles that can guide different speaker's application of a moral expression. Let's consider the concept <vixen> as an example again. According to the classic theory of concepts, we can discover that the proposition 'vixens are female foxes' is true by doing conceptual analysis. A competent speaker, who possess the concept <vixen>, should not be guided by other descriptive principles and thus apply that concept to things that are not female foxes. This is because these types of conceptual truths are very specific. If the conceptual truth realists, who accept Foot's view of moral concepts, really think that moral truths are conceptual truths, then they would need to be able to argue that moral truths too are equally specific. They are required to explain why only one specific descriptive principle, rather than many different plausible descriptive principles, must guide speakers when they use a moral concept like <good>. However, Foot's view on moral concepts cannot really help conceptual truth realist explain this. As a result, even if Foot's view of moral concepts were true, the 2nd premise of the Hare-style argument still seems to be true. This is why Foot's response to Hare cannot help the conceptual truth realists to reject the Hare-style argument. That argument can still be used to argue that the truth of the conceptual truth realism would not be able to accommodate many intuitively plausible cases of moral disagreement.

To summarize, in this §5.4, I have introduced the Hare-style argument against the application of the classical theory of concepts to moral vocabulary. If the argument in question is right, as I have argued, then the classical theory of concepts must be false at least when it comes to moral concepts. This means that the Conceptual Truth Objection must also be rejected because the classical theory of concepts that it relies on turns out to be false. Moral concepts just do not have the kind of substantial descriptive content that could be discovered through conceptual analysis in a way that would be able to avoid the evolutionary debunking argument.

5.5 Reply 3: Why We Would Have Evolved to Have the Competency to Conceptually Analyze Moral Concepts

Finally, §5.5 will introduce a third possible reply to the Conceptual Truth Objection. This reply can be considered to be the denial of the 3rd premise of the objection. That is, I will attempt to reject the premise according to which 'even if the evolutionary forces have significantly influenced and shaped our moral beliefs, we still have the competency to discover which of our moral beliefs are true by doing conceptual analysis' (see §5.2). If this premise turned out to be false, then the Conceptual Truth Objection would be considered to be unsound too. Hence, in this §5.5, I will argue against the idea that we evolved to have reliable capacities to do conceptual analysis in a way that can avoid all the epistemic problems pointed to by the evolutionary debunking argument.

Let us start from considering why the conceptual truth realists think that they have already provided an evolutionary account of why we have the competency to justify our moral beliefs in a reliable way by doing conceptual analysis. It seems that no conceptual truth realists really aim to argue that such a competency would itself be fitness-enhancing. Instead, as mentioned in 5.2.2, they argue that the capacity to justify our moral beliefs by doing conceptual analysis is thought to be an extension or a manifestation of a more general capacity that we have evolved to have – the capacity to come to know conceptual truths in all domains more generally.

Conceptual truth realists rely heavily on the analogy of mathematics to explain why we have that competency to justify our moral beliefs by doing conceptual analysis. According to them, "advanced cognitive skills" in advanced mathematics are not themselves fitness-enhancing (Cuneo & Shafer-Landau 2014: 427), because these skills are just a manifestation of a simpler capacity to do simple mathematics. Furthermore, we can be certain that we have evolved to have the capacity to do simple mathematics because that capacity is clearly fitness-enhancing. For instance, as I previously explained in 5.2.2, the ability to do arithmetic enables us to make correct predictions about dangerous animals or the amount of food remaining. It is just a happy coincidence that we also have the extended advanced cognitive skills to appreciate advanced mathematics as well.

In this reply, I will not attempt to argue that we have not evolved to have the capacity to come to know conceptual truths generally. Instead, I merely aim at arguing that, even if we had the general capacity to come to know many non-moral conceptual truths,

having that capacity would not guarantee that we would be able to justify our moral beliefs by means of conceptual analysis. To argue against this view, I suggest that conceptual truth realists are confused about two important things here. Firstly, they mistakenly assume that, if one has evolved to have a given simpler capacity, this would also entail that one must also have evolved to have the relevant advanced capacity, where that advanced capacity would just be an extension of the simpler capacity. Secondly, we must recognize that morality and mathematics are two very different domains. It might be true that the conceptual truth realists' view of our extended capacity to appreciate advanced mathematics is right. However, this would not necessarily entail that we would also have the extended capacity to justify our moral beliefs in a reliable way by conceptually analyzing moral concepts (even if that we had the capacity to know the conceptual truths in other domains).

Let us start from the conceptual truth realists' first confusion and consider chimpanzees as an example. It seems that chimpanzees can also do simple arithmetic and mathematics even though they are not able to appreciate higher mathematics like topology.[126] For this reason, we should not directly infer that we have evolved to have a higher capacity simply because that capacity could be argued to be just a manifestation or an extension of a relevant and more general capacity that we have evolved to have. The claim that we have the higher capacity needs some independent support other than the fact that we have evolved to have the related simpler capacity.

[126] For representative experiments on chimpanzees about their capacity to do simple arithmetic and mathematics, see Boysen (1993) and Woodruff and Premack (1981).

Conceptual truth realists might agree with me on the previous point. This is because they probably would go on to point out that human beings and chimpanzees belong to different species. We humans, unlike chimpanzees, do have the extended capacity to appreciate higher mathematics – obviously we have such a capacity. However, I also want to argue that it is not merely a happy coincidence or matter of luck that we humans, and not chimpanzees, have the extended capacity to appreciate higher mathematics.

Recently, scientists have shed some light on why a species that has evolved to have a certain simpler capacity need not necessarily have evolved to have the relevant higher capacity too. Michael J. Beran (2008) has argued that not all mathematical capacities are suited to support survival. For instance, recent scientific research shows that chimpanzees are able to tell the difference between "a tree with 10 pieces of fruit from another with only six pieces" because they have a better chance of surviving with this capacity (Beran 2008: 221-222). Yet, the same research also shows that they are struggling to spot the difference between the trees when the difference between the amount of fruits in those trees becomes smaller (Beran 2008: 221). For example, chimpanzees are more likely to fail to distinguish the tree with 5 pieces of fruit from another with 4 pieces.

In light of the results of these experiments, Beran (2008) concludes that it would not have helped chimpanzees to survive as a species if they had evolved to spot the difference between 24 and 28 pieces of fruit (or between 9 and 10 predators) (221). Instead, they only need the capacity of understanding "approximate numerical representations" that is sufficient to support their survival in their natural environment

(Beran 2009: 194). This is why the chimpanzees have only evolved to have the simpler mathematical capacity rather than the capacity to appreciate advanced mathematics. Thus, that one has evolved to have a simpler capacity does not always mean that one must have also evolved to have the relevant higher capacity.

Why do humans then have the capacity to appreciate higher mathematics, including the ability to be able to tell the difference between 24 and 28 pieces of fruits? According to scientific research, this is because the capacity to appreciate higher mathematics is indeed fitness-enhancing in itself. Humans and chimpanzees have been living in very different societies and natural environments, which is why Beran (2009) argues that our capacity to appreciate higher mathematics actually helped our survival in the environments in which we evolved (194). For instance, as he suggests, human children have been raised in the environments where "numerical information is everywhere, and number words and number symbols are used frequently" (Beran 2008: 222). Chimpanzees, by contrast, did not evolve in that kind of natural environments. Hence, it would not have been fitness-enhancing for chimpanzees to acquire that capacity to appreciate higher mathematics.[127]

To sum up, scientific research seems to show that the best explanation of why we have evolved to have the capacity to appreciate higher mathematics is that this capacity itself is fitness-enhancing rather than an extension of the simpler capacity to do simpler

[127] Interestingly, Beran (2008) also suggests that it is possible that non-human animals could evolve to acquire the capacity to appreciate higher mathematics one day given that these animals might be raised in the environments where they would be required to frequently use number symbols and complicated mathematical skills (222).

mathematics. This means that the conceptual truth realists have been confused about how humans evolved to have the higher capacity to appreciate advanced mathematics. If this is the case, then they cannot use the analogy of mathematics to understand how we came to have the cognitive capacity to conceptually analyze moral concepts.

As previously mentioned, I believe that the conceptual truth realists are confused about something else too. We should keep in mind that morality and mathematics are two very different domains. Let us assume, at least for the sake of the argument, that the conceptual truth realists are right and we do have the extended capacity to appreciate higher mathematics because we evolved to have the capacity to do simple mathematics. Nevertheless, as I will argue next, this assumption would not necessarily entail that we would have the extended capacity to justify our moral beliefs in a reliable way by conceptually analyzing our moral concepts, even if we had the capacity for learning conceptual truths in other domains.

According to conceptual truth realism, both conceptual truths and conceptual moral truths are true in virtue of their constituent concepts. However, there actually is a significant difference between moral concepts (such as <good> and <bad>) and many ordinary non-moral concepts (such as <vixen> and <bachelor>). The fact that we have the capacity to come to know conceptual truths about vixens and bachelors itself does not mean that we would have the extended capacity to access conceptual moral truths as well. This is because these capacities are not on the same spectrum. In other words, the capacity to come to access the conceptual moral truths is not a 'higher' capacity, nor is the capacity to access conceptual truths generally in other domains a 'simpler'

capacity. They are very different capacities, and thus the former cannot be considered to be an extension of the latter.

To explain why these two capacities are not on the same spectrum, I rely on a distinction between different kinds of concepts originally introduced by Jerry Fodor (1998). In *Concepts: Where Cognitive Science Went Wrong*, Fodor discusses Hilary Putnam's (1983) article about analyticity (Fodor 1998: 80-85). He suggests that concepts such as <bachelor> and <vixen> can be considered to belong to the same category of concepts, that is, they are "one-criterion concepts" (Fodor 1998: 80). A concept is a one-criterion concept only if there is only one way in which you can apply the concept in question. For example, there is only a way to tell whether something is a vixen or not – by finding out whether it is a female fox (Fodor 1998: 80-81). Furthermore, it seems that conceptual analysis at least in many cases enables speakers to come to know the relevant conceptual truths that are based on the meaning of one-criterion concepts.

By contrast, there are many other concepts that are not one-criterion concepts. These concepts include thin moral concepts such as <right> or <wrong> but also many other philosophically significant concepts, including <free will>, <knowledge>, <justification>, <freedom>, and so on. These concepts can be thought to belong to a different category of concepts, that is, they are what we might call 'philosophically interesting concepts' (Huemer 2015: 51-76).

We can then focus on the key differences between the philosophically interesting concepts and the one-criterion concepts. One of the main differences between these two

categories of concepts is the very different kind of track records we have with respect to analyzing them. The historical track record of conceptually analyzing one-criterion concepts is actually very good. We have long been successful in analyzing one-criterion concepts, including <bachelor> and <vixen>. After all, there isn't much disagreement concerning the definitions of those concepts. In comparison, the historical track record of conceptually analyzing philosophically interesting concepts, such as <knowledge> and <freedom>, is not very good at all.[128]

Let us consider the philosophical interesting concept <knowledge> as an example. Plato is thought to be the first philosopher who attempted to analyze this concept with something like the well-known definition 'justified true beliefs' (*Theaetetus* 210a-b; *Meno* 97a-98b).[129] However, in his article "Is Justified True Belief Knowledge?", Edmund Gettier (1963) casts doubt over this definition of knowledge with his famous counter-examples. Since then, philosophers have suggested numerous analyses of what knowledge consists of, but all these accounts seem to continue to face equally difficult challenges from numerous counter-examples.[130] As Scott Sturgeon (1993) perfectly sums up, a cottage industry, which busily produces responses to Gettier and then new Gettier-style cases against those responses, was born and has had astonishing output

[128] In addition, Fodor (1998) suggests there is a third kind of concepts (81). He calls those concepts – concepts such as <water>, <dog> and <force> – "cluster concepts" (Fodor 1998: 81). Our analyses of the cluster concepts also have a decent track record, compared to the analyses of the philosophically interesting concepts. Nevertheless, the analyses of the cluster concepts are far less reliable than the analyses of one-criterion concepts. For example, the traditional analysis of the concept <force> (i.e., the one according to Newton's laws of motion) is thought to be wrong after the emergence of quantum mechanics and the theory of relativity. However, modern scientists are still able to analyze the concept <force> in a way that can fit our current scientific discoveries.

[129] For example, Edmund Gettier (1963) also suggests that Plato "seems to be considering" and also "accepting" the definition of knowledge as 'justified true beliefs' (121n1).

[130] For summaries of representative attempts to analyze knowledge, see Feldman (2003: 8-38) and Williams (2001: 13-37).

(156). Timothy Williamson (2011) even argues that knowledge is unlikely to be analyzable at all (208-211).

As a result, although many philosophers have attempted to analyze the concept <knowledge>, none of these analyses has really succeeded. Similarly, all the attempts to analyze other philosophically interesting concepts like <free will>, <freedom> or the thin moral concepts seem to have suffered from the same fate. Due to the unsuccessful track record of the analyses of the philosophically interesting concepts, we can arguably conclude that we humans have not really evolved to have the capacity to conceptually analyze the philosophical significant concepts, including the thin moral concepts like <good>, <bad>, <right>, <wrong>, etc. If we had evolved to have such a capacity, it would be a complete mystery of why we would be so bad at analyzing these concepts. Therefore, in light of such a bad track record, it can be argued that, if we formed moral beliefs by conceptually analyzing moral concepts, this belief-formation method would actually be unreliable.

As mentioned in 5.2.2, conceptual truth realists have suggested that we do have the capacity to analyze the moral concepts because it is just a manifestation or an extension of the more general capacity of doing conceptual analysis successfully in other domains. It is true that we have the capacity for coming to know conceptual truths that are based on the meaning of the one-criterion concepts. Yet, it is implausible that this capacity can really help us to analyze moral concepts, which belong to the philosophically interesting concepts. If the capacity for analyzing the latter kind of concepts were actually an extension of the former capacity, then our attempts to analyze the moral

concepts should not have such a bad track record. This shows that the capacity to analyze one-criterion concept like <bachelor> cannot really help us to analyze moral concepts like <good>. Thus, the capacity to analyze one-criterion concepts does not really extend to the philosophically interesting moral concepts.

Conceptual truth realists might want to argue at this point that the analyses of at least some moral concepts actually have a decent track record. However, if they pursued that line of response, it is likely that they would fail to provide any convincing examples. We can look at the examples of the conceptual moral truths suggested by the conceptual truth realists critically. As mentioned in 5.2.1, they suggest that moral propositions such as 'it is right to maximize happiness' or 'it is wrong to kill a child merely for fun' are core examples of conceptual moral truths. Let us begin from the moral belief 'it is right to maximize happiness' as it is already a very controversial moral proposition. Although many philosophers defend it, there are a number of plausible opponents who reject this proposition. Deontologists reject this. Virtue ethicists reject this. This moral proposition is thus not convincing enough to be considered to be a successful example of analyzing the moral concept <right>.[131]

Yet, it could be suggested that the moral belief 'it is wrong to kill a child merely for fun' is a more convincing example of genuine justified moral beliefs. However, even if this

[131] Interestingly, the moral belief that it is right to maximize happiness could still arguably enhance our fitness. For example, Charles Darwin (1871/2009), the founder of the evolutionary theory, accepted the utilitarian "greatest happiness principle" in his book *The Descent of Man and Selection in Relation to Sex* (97-98). He argued that we should maximize the greatest general good and welfare, and the "foundation of morality" should also lie in the utilitarian-style of "greatest general good principle" (Darwin 1871/2009, 97-98). Nevertheless, even if Darwin were right, this would only mean that we evolved to have the moral belief that it is right to maximize happiness merely because of the evolutionary influence on our moral beliefs (see also Chapter 2 of this thesis) and not because of conceptual analysis.

particular moral belief were justified, this would not necessarily imply that we thus would have the good track with respect to analyzing the moral concept <wrong>. Let's consider the philosophically interesting concept <knowledge> as an example again. Most of us, with the exception of the sceptics, accept that we do have a wide range of knowledge. Nevertheless, the fact that we have a wide range of knowledge does not entail that we have succeeded in analyzing the concept <knowledge>. Otherwise, the Gettier problem would be easily solved. Hence, likewise, even if the moral belief 'it is wrong to kill a child merely for fun' were justified, this would not entail that we thus have a good track record of analyzing moral concepts, including the concept <wrong>. After all, showing that we have the ability to analyze moral concepts successfully requires much more than this particular moral belief.

As a result, we should be sceptical about the idea that we have the competency to analyze moral concepts – our track record when it comes to the conceptual analyses of those concepts just is too bad for that claim to be plausible. Moreover, if we have reasons to be skeptical about that, then it is not plausible to argue that we could justify our moral beliefs in a reliable way by conceptually analyzing them. Likewise, it is not plausible to think that the capacity to come to know conceptual truths in other domains could help us to have that capacity to justify those moral beliefs by conceptual analysis.

To sum up, the conceptual truth realists are required to provide an account of how we evolved to have the competency to conceptually analyze moral concepts. This would enable them to argue that our moral beliefs can be justified in this reliable way that is not threatened by any concerns about the evolutionary origin of these moral beliefs.

There are two potential explanations available to them. Firstly, they could argue that the competency is itself fitness-enhancing. But it would be very difficult to argue for this thesis, and it seems that conceptual truth realists have so far not attempted to argue in this way. The second way is to argue that the capacity is an extension of a 'simpler' capacity to conceptually analyze concepts in other domains. However, in this §5.5, I have also explained why the capacity to conceptually analyze moral concepts should not be considered to be an extension of the capacity to conceptually analyze concepts in other domains. I can thus conclude that it is unlikely that our moral beliefs can be justified by relying on conceptual analysis because we have a bad track record of analyzing philosophically interesting concepts, including moral concepts.

5.6 Conclusion

In this Chapter 5, I introduced and also rejected the conceptual truth realists' objection to the EDA. Firstly, in §5.2, I explained the basic crux of the Conceptual Truth Objection that relies on the claim that moral truths are conceptual truths. Secondly, in §5.3, I explained why Evers and Streumer's objection against the moral fixed points theory only partially works as a reply to the Conceptual Truth Objection. Then, in §5.4, I introduced the Hare-styled argument against the view that the classical theory of concepts is also applicable to moral vocabulary. If that argument is right, as I have argued, this also means that the Conceptual Truth Objection should be rejected because the conceptual truth realists have essentially relied on that theory. Finally, in §5.5, I suggested that the conceptual truth realists also fail to provide a compelling reason to believe that our moral beliefs could really be justified by conceptually analyzing the moral concepts. All things considered, I then conclude that my objections in §5.4 and

§5.5 show that the Conceptual Truth Objection fails as an objection to the EDA.

Chapter 6

The Third Factor Objection

6.1 Introduction

In this Chapter 6, I will focus on the second of two strongest realist objections to the EDA. This second objection is the 'Third Factor Objection' (Copp 2008; Enoch 2010, 2011; Wielenberg 2010, 2014, 2016). As I already explained in Chapter 2, in order to create an epistemological challenge for the moral realists, the evolutionary debunkers start from the evolutionary origin of our moral beliefs. The defenders of the Third Factor Objection, who are all realists, concede that the evolutionary forces have had a significant influence on our moral beliefs. However, they reject the debunkers' premise that, when evolution influences our moral beliefs, it influences those beliefs in a way that would make them not track the moral truth reliably. This is because, on their view, there are certain additional factors (hereafter the third factors) that can ensure that there is an indirect and yet reliable correlation between our moral beliefs and the moral truth. Hence, according to the Third Factor Objection, even if the debunkers could wholly explain the origin of our moral beliefs in evolutionary terms, the evolutionary origin of those beliefs is insufficient to undermine the epistemic status of those beliefs.

The main objective of this Chapter 6 is to defend the EDA against the Third Factor Objection. This chapter has five parts. Firstly, §6.2 will introduce the Third Factor Objection and explain how it could be used to argue against the EDA. Then, in §6.3-5. I will introduce three dominant versions of the Third Factor Objection suggested by

David Copp (2008), David Enoch (2010; 2011) and Erik Wielenberg (2010, 2014, 2016) respectively. Finally, in §6.6, I will argue that all these three versions are too problematic for three different reasons. In that sub-section, I will eventually conclude that the defenders of the Third Factor Objection will not be able to formulate a plausible version of the Third Factor Objection because it is unlikely that they could formulate a version that could avoid the three problems of the versions explored in this chapter.

6.2 The Objection

In this §6.2, I will introduce the Third Factor Objection and explain how this objection could be used to argue against the EDA. As mentioned in §6.1, the defenders of the Third Factor Objection also accept the evolutionary origin of our moral beliefs. Nevertheless, they want to argue that the fact that the evolutionary forces have influenced our moral beliefs need not itself necessarily mean that our moral beliefs would not be tracking the moral truth reliably. Indeed, they too concede that they cannot provide an explanation of how those moral beliefs could be directly aligning with the corresponding moral truth. However, they instead provide third factor explanations which relate our moral beliefs to the moral truth by relying on a third factor where this third factor is correlated with both our moral beliefs and the moral truths.

To further explain how the third factor explanations work, let us consider the difference between a direct correlation and a third factor correlation. By a direct correlation between A-factors and B-factors, David Enoch (2011), who is a defender of the Third Factor Objection, means that the A-factors are somehow causally or constitutively responsible for the B-factors, or vice versa (167). Therefore, if we are required to

explain how the moral truth could be directly correlated with our moral beliefs, then we need to explain how the moral truths themselves could be related to or responsible for our moral beliefs, or vice versa. In the previous Chapter 3, I introduced two views of how moral truth could be directly correlated with our moral beliefs – the liberal version of the explanatory reading and the modal reading of truth-trackingness. In that chapter 3, I also concluded that there is no such a direct correlation between our moral beliefs and the moral truth due the evolutionary origin of those beliefs, no matter which one of the two readings of truth-trackingness we accept.

In contrast, according to the third factor explanations, two factors A and B can still be aligned without being causally or constitutively responsible for each other. This is because they can be aligned in an indirect way when there is a third factor C that is responsible for both factors A and B (Enoch 2011: 167). Hence, according to the Third Factor Objection, if there is such a third factor which is responsible both for our moral beliefs and for moral truths, it can also explain why our moral beliefs could be aligning with the moral truth in an indirect yet reliable way even if those beliefs are not directly correlated with the corresponding truth (Wielenberg 2016: 505). If the Third Factor Objection were right, then the epistemic status of our moral beliefs would not be undermined and the EDA could thus be rejected.

But how could our moral beliefs correspond to the moral truth in such an indirect way because of the third factor? Let us consider the following Figure 6.1, which illustrates the basic crux of the third factor explanations:

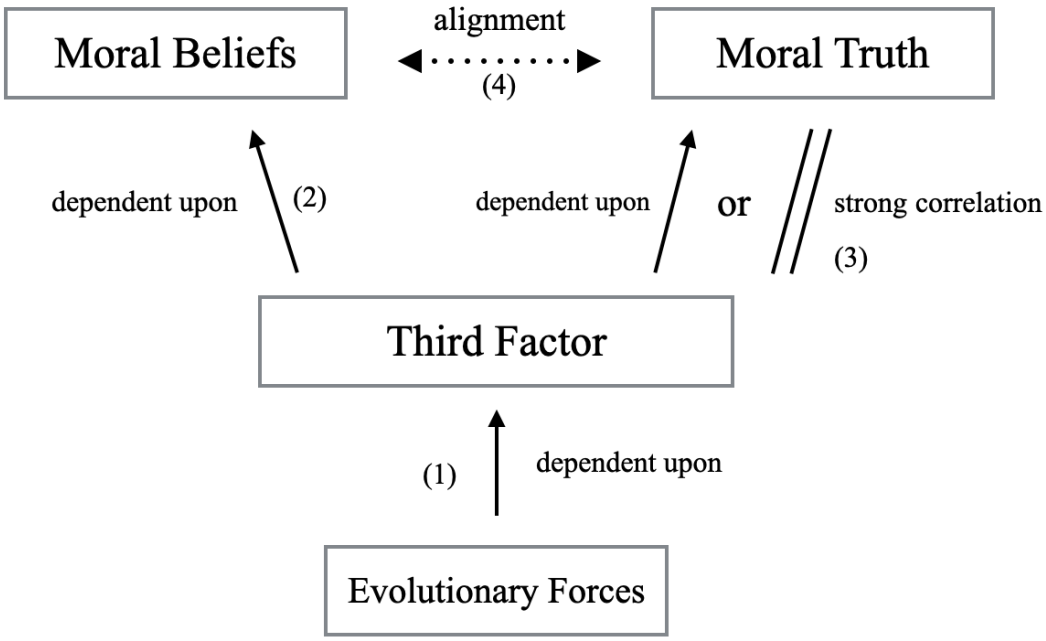

Figure 6.1 – The basic crux of the third factor explanations

All versions of the Third Factor Objection hold that there are four kinds of relations, which are indicated in Figure 6.1, in the third factor explanations. In the rest of this sub-section §6.2, I will explain those four kinds of relations respectively:

(1) The relation between the third factor and the evolutionary forces (6.2.1)

(2) The relation between the third factor and the moral beliefs (6.2.2)

(3) The relation between the third factor and the moral truth (6.2.3)

(4) The relation between the moral beliefs and the moral truth (6.2.4)

Although the structure of the different versions of the Third Factor Objection is basically the same, different versions explain the four relations above very differently. In the rest of this §6.2, I will explain the questions of which the defenders of the Third

Factor Objection need to answer when they explain these four relations.

6.2.1 The Relation between the Third Factor and the Evolutionary Forces

Let's start with considering the relation, which is denoted with (1) in Figure 6.1, between the third factor and the evolutionary forces. In general, the third factor explanations suggest that the third factor is dependent upon the evolutionary forces. Hence, I have denoted the kind of dependence relations between the third factor and the evolutionary forces suggested by the defenders of the third factor explanations – including Wielenberg, Copp and Enoch – with an arrow (➔) in Figure 6.1.

Nevertheless, why would there be a dependence relation between the third factor and the evolutionary forces? Different versions of the third factor explanations provide different explanations of the dependence relation depending on what, according to their version, the third factor is. Therefore, different versions of Third Factor Objection differ from one another with respect to exactly what their answer is to the following the first essential question:

- Question TF1: What is the third factor?

I will further explain their answers to this question and their descriptions of the correlations between the third factor and the evolutionary forces later in Sections §6.3-5.

6.2.2 The Relation between the Third Factor and the Moral Beliefs

We can then consider the relation, which is denoted with (2) in Figure 6.1, between the third factor and our moral beliefs. All versions of the third factor explanations basically share the view that our moral beliefs have been formed under a significant influence of the third factor. Hence, the defenders of the third factor explanations thus argue that our moral beliefs are dependent upon the third factor, and I have denoted the kind of dependence relations between the third factor and our moral beliefs with an arrow (➔) in Figure 6.1. Moreover, as we saw above, the third factor too is dependent upon the evolutionary forces. As a result, the third factor explanations are compatible with the view that the evolutionary forces have significantly influenced and shaped our moral beliefs (via the third factor).

To further explain the dependence relation between the third factor and our moral beliefs, we can consider the example of an earthquake and a pill. Imagine that I took a pill just before I formed the belief that an earthquake has just happened in the city X. Let us further assume that in this case the fact that I form the previous belief is a causal consequence of taking that pill. Thus, the formation of my previous belief is dependent upon the pill that I took. Analogically, we can then compare the third factor to the pill that I took and our moral beliefs to my belief that an earthquake has just happened in the city X. This means that our moral beliefs are dependent upon the third factor in the same way in which the formation of the belief that an earthquake has just occurred in City X is dependent upon the pill that I took.

Nevertheless, for the purposes of their explanations of the dependence relation between

our moral beliefs and the third factor, *all* versions of Third Factor Objection need to make certain assumptions on the connection between the third factor and our moral beliefs. As previously mentioned, the defenders of Third Factor Objection also accept the evolutionary origin of our moral beliefs. This means that they need to make some claims concerning just how our moral beliefs came to be correlated with the relevant third factor, which was influenced by the evolutionary forces. Therefore, the second essential question which the different versions of Third Factor Objection need to answer in different ways is:

- Question TF2: What claims do the defenders of the Third Factor Objection make about the evolutionary origins of the third factor and the way it is responsible for our moral beliefs?

Again, I will further introduce their answers to this question and outline their descriptions of the relation between the third factor and our moral beliefs in §6.3-5.

6.2.3 The Relation between the Third Factor and the Moral Truth

Let us then consider the relation, which is denoted with (3) in Figure 6.1, between the third factor and the moral truth next. There are two possible ways to explain the correlation between the third factor and the moral truth. Firstly, Copp (2008) and Wielenberg (2014) think that moral truth is dependent upon the third factor. To explain this kind of a dependence relation, let us consider the example of the earthquake and the pill again. Let us further imagine that the pill I took was not a usual pill, but rather it was created by a powerful witch. The pill thus had two effects. Firstly, as previously

mentioned, I formed the belief that an earthquake has just occurred in city X entirely because of the pill. However, the other effect of the pill is that, whenever someone takes such a pill, there really is an earthquake in city X. As a consequence, the fact that an earthquake has just occurred in city X always correlates with the beliefs of the person taking the pill.

Analogically, we can compare the third factor to the pill again and also the fact that an earthquake has just happened in city X to moral truth. Just like the example of the earthquake and the pill, the defenders of the Third Factor Objection suggest that the third factor is not merely correlated with our moral beliefs but also related to the moral truth in some way. Unlike the example in question, however, they do not aim at arguing that there must be a causal relation between the third factor and the moral truth. Instead, as we will see in this Chapter 6 later on, they argue that moral truth can be understood to be dependent upon or responsible for the third factor. As a result, I have illustrated these dependence kinds of correlations between the third factor and the moral truth with an arrow (➔) in Figure 6.1.

As we will see below, David Enoch (2011) explains the relation between the third factor and the moral truth in a different way. He does not think that that relation is a dependence relation. Instead, he argues that there is a coherence type of a relation between the relevant third factor and at least some moral truths (Enoch 2011: 169). I have illustrated this kind of a coherence correlation with a double hyphen (=) in Figure 6.1.

Again, for the purposes of their third factor explanations of the connection between the third factor and moral truth, *all* versions of Third Factor Objection need to make certain assumptions. More precisely, the defenders of the Third Factor Objection need to provide an explanation of just why and how the third factor is supposed to correlate with the moral truth. Whilst providing such an explanation, they must rely on at least some views concerning what the moral truths are. These views are required because we can determine whether the third factor is correlated with the moral truth only if we have at least some idea of what the moral truth (and the third factor) is. Therefore, the third essential question that the different versions of Third Factor Objection need to answer in different ways is:

- Question TF3: What views do the defenders of the Third Factor Objection hold about the moral truth and the way in which the third factor is related to it?

Different defenders of the third factor explanations hold different views concerning the connection between the third factor and the moral truth. I will further describe their views of the relation between the third factor and the moral truth in §6.3-5.

6.2.4 The Relation between the Moral Beliefs and the Moral Truth

Finally, we can consider the relation between our moral beliefs and the moral truth, which is denoted with (4) in Figure 6.1. To explain this relation in the third factor explanations, we can again recall the example of the magical pill, the earthquake and the witch. In that example, there would be an indirect connection between my belief (an earthquake has just occurred in city X) and the corresponding fact (an earthquake

has actually just occurred in city X) via the agency of the witch. In that case, both my belief and the corresponding fact would depend on that magical pill and so, as a result, my belief that an earthquake has just occurred in city X would not be accidentally true. Thus, there would, in this situation, be an indirect and yet reliable correlation between the belief and the corresponding fact. As a consequence, the epistemic status of my belief concerning the earthquake would not be undermined even if there were no direct causal relation between my belief and the corresponding fact.

Following the same line of reasoning, the third factor explanations do not aim at explaining the correlation between our moral beliefs and moral truth directly (for example, by relying on something like a direct causal connection between our moral beliefs and the moral truth). Instead, they merely aim at providing an explanation of the correlation between our moral beliefs and moral truth in a more indirect, mediated sense. According to the Third Factor Objection, if the third factor is able to ensure such an indirect but reliable correlation between the moral truth and our moral beliefs, it is sufficient to ensure that our moral beliefs have been attained in a reliable way. Thus, the epistemic status of these beliefs would not be undermined (Crow 2016: 389).

Nevertheless, there is another critical question that the defenders of the third factor explanations need to answer. Let us assume that there is a third factor which is responsible for both our moral beliefs and the moral truth. Moreover, let us also assume that this third factor can explain why there is a minimal, indirect correlation between our moral beliefs and the moral truth. Yet, even in this situation we can further ask: How can the third factor explanation ensure that our moral beliefs can be attained in a

reliable way? Even if the explanations in question could ensure that there is a minimal correlation between our moral beliefs and moral truth, does this really entail that these moral beliefs have been attained in a sufficiently reliable way so that the epistemic status of these beliefs would not be undermined as a consequence of their evolutionary origin? Thus, the fourth essential question that the defenders of the Third Factor Objection need to answer is:

- Question TF4: Why is the third factor explanation sufficient for enabling us to reject the EDA (In other words, why is this explanation sufficient to ensure that the epistemic status of our moral beliefs would not be undermined)?

Similar to the three essential questions mentioned above, different versions of Third Factor Objection differ from one another with respect to exactly what their answer to this fourth question. And, I will further explain their answers to this question in the coming §6.3-5.

6.3 David Copp's Justified Social Moral Codes

I will then introduce three dominant versions of the Third Factor Objection next. These have been suggested by David Copp (2008), David Enoch (2010; 2011), and Erik Wielenberg (2010; 2014; 2016) respectively. I will introduce these views as responses to the questions TF1-4 explained in the previous section. In the remaining part of this §6.3, I will focus on Copp's view. Then, in the next Sections §6.4 and §6.5, I will turn to how Wielenberg and Enoch answers to the questions introduced in the previous section.

6.3.1 Copp's Reply to Question TF1

What is the third factor according to David Copp's version of the Third Factor Objection? In his article "Darwinian Skepticism about Moral Realism", Copp (2008) suggests that justified social moral codes are the third factor that is responsible for both moral truth and our moral beliefs. A moral code, by definition, is a system of moral standards and norms that can be internalized by a person or a society (Copp 2001: 5-6). If a moral code is internalized by a significant majority of a society, then this moral code can be considered to be the social moral code of that society. Let us consider the Ten Commandments as an example. The Ten Commandments can be considered to be a moral code, which consists of ten moral principles. Moreover, most people in Christian societies have endorsed those ten moral principles – such as thou shalt not kill and thou shalt not steal. As a result, the Ten Commandments, which are internalized by most people in Christian societies, thus constitute the social moral code of those Christian societies.

Moreover, according to Copp (2001), a social moral code can be justified for a given society only if most members of that society could have rationally selected that moral code (104). For instance, let us consider a society the members of which could rationally select the Ten Commandments because those rules satisfy their needs and values.[132] This would also entail that the members of that society would prefer to

[132] As Copp (2001) suggests, there are also other instrumental theories of rational choice, such as the "welfare theory" and the "desire theory" (169). According to those theories, members of a given society would rationally select the social moral code that would best maximize their welfare (according to the welfare theory) or the satisfaction of their desires (according to the desire theory). However, Copp (2001) rejects both these theories because he thinks that a rational person would prioritize her values and own

endorse the Ten Commandments rather than any other moral code after rational reflection in which they would consider which code would best serve their needs and values. If this were the case, the Ten Commandments could be thought to be a justified social moral code for the society on Copp's view.

We can then consider how Copp understands the dependence relation between the evolutionary forces and the third factor. On his view, the third factor is dependent upon the evolutionary forces because we have evolved to have justified social moral codes. In Chapter 2, I already explained how we have evolved to have different traits. Briefly, an evolved (or adapted) trait for a species must have two essential features. Firstly, the evolved trait must serve a certain purpose or a function for them in a teleological sense. Secondly, having such a trait with that function enables this species to leave more offspring in the circumstances that they have been living in.

Based on what I introduced in Chapter 2, it can be argued that we humans evolved to have justified social moral codes as the general internalization of these codes also has the two essential features of the evolved traits. As I previously explained, a social moral code is justified for a given society if it can best meet the needs and values of the members of the society (Copp 2008: 200-201). Hence, by definition, a justified social code must have the first feature of an evolved trait – it must serve a function in a teleological sense to meet these basic needs.

We can imagine a society P with a justified social moral code that can enable all the

needs ahead of her welfares or desires (169).

members of the society to meet all their basic needs and also another society Q, which is in a similar circumstance with P, but without any justified social moral code. Presumably, in this situation, the society P's justified social moral code would promote behaviour that meets the basic needs of its members and also crucially increase their chance of survival (Copp 2008: 201). For example, all societies arguably have the basic need to maintain peacefulness (Copp 2008: 200-201). Hence, it is safe to assume that society P's justified social moral code too would promote behaviour that encourages avoiding violence amongst its members. Moreover, it is also reasonable to think that a society that can maintain peace and avoid violence among its members would arguably have a greater chance to leave more offspring in the succeeding generations.

In contrast, the members of the society Q in a similar circumstance would find it more difficult to maintain peace and co-operation due to the lack of a justified social moral code in their society. Hence, the members of the society Q would eventually leave fewer offspring in succeeding generations than those of the society P. As a result, the justified social moral code can be thought to have the second feature of an evolved trait because having such a code enabled us to leave more offspring in the circumstances we were living in a long time ago. According to Copp, it is thus very likely that we have evolved to internalize a justified social moral code.

6.3.2 Copp's Reply to Question TF2

In order to explain how a justified social moral code could be the third factor connecting moral truth and our moral beliefs, Copp makes certain claims about the evolutionary origin of our moral beliefs and also the third factor. As we just saw above, he suggests

that we have evolved to internalize a justified social moral code. When a society internalized a justified social moral code under the evolutionary influence, the internalization also requires the members of that society to subscribe to that justified social moral code (Copp 2008: 201). Moreover, if they subscribed to the justified social moral code, then they would tend to form certain moral beliefs that correspond to the moral standards included or implied by that moral code (Copp 2008: 205n18). In previous 2.2.2, I mentioned that our moral beliefs are reliably connected to what we are motivated to do. As a result, since we internalized a justified social moral code under the evolutionary influence and thus formed moral beliefs that corresponded to that code, we also came to act in a way that enabled us to leave more offspring.

Let us consider a social moral code which includes the moral standards that promote co-operation as an example. However, this code could have enabled us to leave more offspring only if we were able to act according to it (see also §2.3). We can further assume that our society internalized this social moral code due to the evolutionary influence. This assumption means that the former members of our society subscribed to that code and also tended to form moral beliefs such as that it is good to co-operate with others (Copp 2008: 201-202). Furthermore, our ancestors would have also had the corresponding motivation to act according to that moral belief. As a consequence, they were motivated to act in a way that corresponded to the moral standard that promoted co-operation, which is included in the social moral code in question (Copp 2008:202).

To sum up, according to Copp, if we have evolved to internalize a justified social moral code, then we will also have evolved to have the moral beliefs that correspond to that

justified social moral code as well. Our moral beliefs are dependent upon the justified social moral code (i.e., the third factor) in the way described above on his view.

6.3.3 Copp's Reply to Question TF3

In order to explain how justified social moral codes could also be the kind of a third factor that is not only responsible for our moral beliefs but also for the moral truth in some sense, Copp relies on a certain normative claim – there is a grounding relation between the justified social moral code and the moral truth. To be more precise, he thinks that the moral norms and standards included in or implied by the justified social moral code have a "truth-grounding status" (Copp: 2008: 199). For instance, on his view, the moral proposition that wanton killing is wrong is true if and only if and just because wanton killing is forbidden by the norms and standards that are included in the relevant justified social moral code (Copp 2001: 23; 2008: 205n18).

Just how could the justified social moral code ground the moral truth? Copp's response to this question begins from the idea that moral truths, as a kind of normative truths, must have authority, or what is sometimes called normativity. Copp (2001) then argues that the authority or normativity of the moral truths can only be "explained in terms of motivation or reasons" (32). More precisely, he means that moral truths have authority or normativity only if we have reasons and motivations to comply with those truths. According to Copp (2001), a given social moral code can have authority over us only if we take the moral standard included in that code to be true (25). Furthermore, on his view, we take a given moral standard to be true if and only if two further conditions are both satisfied. Firstly, we need to have at least some motivation to act according to that

moral standard. Secondly, we also must have some justification for the moral standard. Therefore, when we take a moral standard to be true, this also means that this moral standard will have the authority over us because we must have reasons and motivation to comply with it.

We can then consider why the relevant justified social moral codes would have authority. As mentioned in 6.3.1, a given social moral code is justified for us only if we would choose to adopt that moral code after rational deliberation or reflection. Moreover, if we could rationally decide to subscribe to a given social moral code, then we would also be motivated to act according to the moral standards included in that social moral code. This is because being able to rationally choose to adopt a social moral code also requires having motivated reasons to comply with the moral standards included in that social moral code on the basis of needs and values. Hence, the moral standards included in a justified social moral code must have authority over us because we must have at least some motivation and reasons to comply with them after rational deliberation. Furthermore, as mentioned above, moral standards can be true only if they also have authority (Copp 2008: 199). As a result, the moral standards which are included in a justified social moral code must also have a truth-grounding status because moral standards of such kind must have authority.

Let us consider the moral proposition that wanton killing is wrong as an example again. Let's also imagine that the justified social moral code of a society Y forbids the acts of wanton killing. According to Copp, the moral proposition that wanton killing is wrong is true in the case above because for him the justified moral code of a society grounds

the moral truths of that society. As he puts it, the moral proposition in question is true if and only if that wanton killing is forbidden by the standard that is included in a justified social moral code (Copp 2001: 24). If the social moral code including the moral standard that forbids the action of wanton killing is justified for the society Y, then its members must also be motivated and have reasons to comply with that moral standard because they could choose to accept the code after rational deliberation or reflection (Copp 2001: 34). And, this is guaranteed by the fact that a justified social moral code must help the members of the society to satisfy their needs and values. This means that the moral proposition that killing is wrong can be considered to be true in virtue of the authority of the previous moral standard included in the justified social moral code. As a result, when Copp suggests that the justified social moral code (i.e., the third factor) can ground the relevant moral truth, he actually means that the authority of the related moral truths is dependent upon the justification for the relevant code that is based on the way in which the justified social moral code helps the members of the society to satisfy their needs and values – something they are motivated to do.

6.3.4 Copp's Reply to Question TF4

Let's return to how the EDA argues that the epistemic status of our moral beliefs is undermined. As mentioned in the previous Chapters 3 and 4, according to the evolutionary debunkers, our beliefs are not tracking the moral truth because these beliefs have not been formed in a way that would make them adequately align with the corresponding truths. Hence, given that our moral beliefs do not adequately align with the corresponding moral truths due to their evolutionary origin, the epistemic status of these beliefs is undermined.

In response to the EDA, Copp argues that evolutionary debunkers are mistaken to argue that our moral beliefs are formed in a way that makes them not aligned with the corresponding moral truths. Based on Copp's answers to TF1-4 we can clearly see why there would be an indirect relation between our moral beliefs and the moral truth as a consequence of the third factor. Firstly, under the evolutionary influence, we humans formed moral beliefs that corresponded to the justified social moral codes. Secondly, those moral beliefs are likely to be true given that the justified social codes that they correspond to are also argued to ground the moral truth. Hence, Copp (2008) argues that our moral beliefs are still able to "quasi-track" the moral truth through the third factor, that is, the justified social moral codes (194). By this he means that, even if our moral beliefs are not directly corresponding to the moral truth, they are still correlated with the moral truth up to an "epistemically sufficient degree" (Copp 2008: 194). He also argues that, because our moral beliefs are correlated with the moral truth up to the epistemically sufficient degree, the Third Factor Objection is sufficient to avoid the sceptical worries resulting from the EDA (Copp 2008: 196).

However, what does the claim that our moral beliefs are correlated with the moral truth up to an "epistemically sufficient degree" actually mean? Copp does not provide an explicit answer to this question. Yet, he does emphasize that it would be an epistemological problem if it would be a "fortunate accident that there is a tendency for our moral beliefs to approximate to the truth" (Copp 2008:198). Moreover, he also suggests that, if our moral beliefs quasi-tracked the moral truth, it would not be a mere coincidence for our moral beliefs to be true (Copp 2008: 197-98).

Based on Copp's claims above, I believe that Copp's version of the Third Factor Objection can be thought of as the following view: When the evolutionary forces influenced our moral beliefs, they influenced our moral beliefs in a way that enabled these beliefs to quasi-track the moral truth via the justified social codes and therefore not to be merely true accidently. Copp seems to think that, if the justified social moral codes (i.e., the third factor) can ensure in the way that our moral beliefs are not merely accidentally true, then the epistemic status of our moral beliefs is not undermined. Unfortunately, he did not further explain why this would be the case.[133] As we will see in the next §6.4, Erik Wielenberg is much more careful when he provides the explanation required for the Third Factor Objection at this point.

6.4 Erik Wielenberg's Cognitive Faculties

Let us then consider Erik Wielenberg's answers to the previous four questions TF1-4 that constitute his version of Third Factor Objection.

6.4.1 Wielenberg's Reply to Question TF1

According to Wielenberg, certain cognitive faculties that we human beings have are the third factor that is responsible for both moral truth and our moral beliefs (Wielenberg 2010: 459; 2014: 145). He starts from the idea that we human beings have certain moral rights, which themselves are a kind of moral facts that make certain moral propositions

[133] Matt Bedke (2014) also suggests that there is an epistemological problem given that it would be a lucky coincidence for our moral beliefs to be true. According to him, the presence of coincidence would make "adequate alignment between belief and fact too coincidental to accept" (Bedke 2014: 114). In Chapters 3 and 4, I explained how the epistemic status of our moral beliefs is undermined as a result of the fact that there is a lack of adequate alignment between them and the moral truths.

true (Wielenberg 2010: 451-452; 2014:145). He then observes that we have certain important and special cognitive faculties that both (i) are required for us to be able to recognize that we have moral rights, such as the right not to be killed and not to be tortured and (ii) ground these rights in the first place.

Let us take our moral right not to be killed as an example. In order for us to be able to think that we have this moral right, we must, for example, be able to understand the consequence of killing and how bad it is for us to lose our life. For this reason, in order for us to be able to think that we have a right not to be killed, we already need the cognitive capabilities required for reasoning, agency, language, social co-operation, grasping the ideas of death and future, and so on. If we have these cognitive faculties, then we are able to claim our right when being threatened. Furthermore, the very same cognitive faculties also lead us to believe that we do have those moral rights. In addition, Wielenberg (2014) also thinks that we human beings have the right not to be killed because of the value of these cognitive faculties – that right is supposed to be grounded in the requirement to protect beings who have such faculties (144-145) (I will further explain this view in the 6.4.3 below). Hence, our certain cognitive capacities are considered to be the relevant third factor in Wielenberg's third factor explanations.

Finally, the previous cognitive faculties are also the "products of evolution" because having them enabled our ancestors to leave more offspring in the circumstances they lived in (Wielenberg 2010: 449). In 5.2.2, I already explained why some of our basic cognitive faculties, such as the general power of reasoning and the ability to analyze concepts, are fitness-enhancing. Let us also consider the cognitive capacity required for

language as a further example here. Our capacity to use language actually helped our ancestors' survival because human language can facilitate co-operation and even promote mutual reciprocity (Smith 2010; Adornetti 2015). As a result, the capacity to use language helped us to leave more offspring, and it can thus be considered to be an evolved trait as well. Since many cognitive faculties that we have can all be considered to be evolved traits, this also means that these cognitive faculties (i.e., the third factor) are dependent upon the evolutionary forces (Wielenberg 2010: 449; 2014: 145).

6.4.2 Wielenberg's Reply to Question TF2

In order to explain how our cognitive faculties could be the third factor connecting moral truth and our moral beliefs, Wielenberg makes an evolutionary thesis about the origin of our moral beliefs and the third factor. He argues that we can provide an adequate evolutionary explanation of our moral beliefs about moral rights (Wielenberg 2014: 137). More precisely, he thinks that the existence of certain cognitive capacities is responsible, via an evolutionary process, for our beliefs about the moral rights.

We human beings commonly believe there are some actions that other people ought not to do to us. For instance, we believe that others simply ought not rape us, enslave us, kill us for fun, and so on (Wielenberg 2014: 138). We can then consider our belief that we have the moral right not to be killed as an example. Why do we humans tend to form the belief that we have this right? According to Wielenberg (2014), this is because the relevant cognitive faculties described above that we evolved to have lead us to believe that we do have this right (145). We tend to form moral beliefs that are evolutionarily beneficial with those cognitive faculties, and the belief that we have a moral right not

to be killed is arguably such a belief. If our ancestors all had this belief and also tried to protect each other's moral right not to be killed, their overall fitness was obviously enhanced. Since most, if not all, moral beliefs about moral rights are evolutionarily beneficial to us, we tend to form the moral beliefs about our possession of moral rights under the evolutionary influence.[134]

6.4.3 Wielenberg's Reply to Question TF3

As mentioned in 6.4.1, we need certain cognitive faculties (for example, the ability to take part in social co-operation and to use language, and so on) to know and grasp the idea that we have the rights not to be raped, enslaved or killed (Wielenberg 2014: 137-38). Furthermore, according to Wielenberg (2014), these cognitive faculties are not only required for being able to have beliefs about the previous rights, but the presence of our moral rights is also guaranteed by the presence of those cognitive faculties (145).

We can start from the core idea of human rights. Essentially, human rights are the rights that "we have simply in virtue of being human" (Griffin 2008: 2). We, as human beings, must have certain distinctive human status. Since the distinctive human status makes us so unique and valuable, it deserves protection and so human rights exist in order to protect this distinctive status (Griffin 2008: 2). Nevertheless, different philosophers disagree on what the distinctive human status is based. For instance, Kantian ethicists suggest that rationality is essential to humanity, whereas James Griffin (2008) argues

[134] Wielenberg (2014) admits that he does not aim to provide an evolutionary explanation of all of our moral beliefs because he thinks that such an explanation is unlikely to be a plausible one (137). Hence, in his book *Robust Ethics*, he merely focuses on providing the evolutionary explanations of some of our moral beliefs, such as our moral beliefs about moral rights (Wielenberg 2014: 137, 149).

that the distinctive human status is thought to be our "normative agency" (2). As for Wielenberg, he argues that our cognitive faculties are of the utmost importance, which is why they deserve protection by human rights.

Why would our cognitive faculties be so essential for us as to deserve moral protection? Firstly, our cognitive faculties enable us to be very sophisticated beings and also to pursue very different kinds of valuable life projects. Without our cognitive faculties, we wouldn't be able to do philosophy, create art, make great scientific discoveries, and so on. The presence of our cognitive faculties is essential to many of the activities and achievements that are valuable and unique to human beings. Moreover, the relevant cognitive faculties – including the ability to use language, the ability to understand moral claims, the rational agency, and so on – are thought to be very advanced capacities. In contrast, worms, for example, do not possess these advanced faculties, which is why they are unable to perform the previous activities. Hence, the possession of the relevant advanced cognitive faculties can be considered to be both essential for human beings and also unique and valuable.

As we humans are so unique and valuable because of our cognitive faculties, we are also worth protecting by the relevant moral rights, including the moral rights not to be killed and tortured. According to Wielenberg (2010), our moral rights are thus grounded in our cognitive faculties in the sense described above (459n61). By making the normative thesis that moral rights are grounded in the cognitive faculties, Wielenberg aims to ensure that there is a reliable correlation between a significant number of moral rights (which themselves are a kind of moral truths) and the third factor – the relevant

cognitive faculties.[135]

6.4.4 Wielenberg's Reply to Question TF4

Just like Copp, Wielenberg (2014) too starts by arguing that his third factor explanations can ensure that it is not a coincidence that our moral beliefs are true in general (153). He also suggests that it would not be a coincidence for our moral beliefs to be true even if these beliefs could not be explained by directly appealing to moral truth (Wielenberg 2014: 153). This is because the correlation between moral truths and our moral beliefs is ensured and explained by the presence of the third factor.

As I explained above, Wielenberg (2010) thinks that certain cognitive faculties that we have evolved to have are responsible for generating our beliefs about the moral rights we have, such as our beliefs about moral rights not to be killed or to be tortured (449-450). As we just saw, according to him, these very same cognitive faculties also guarantee the presence of the rights of which those beliefs are about. This means that, according to Wielenberg, these cognitive faculties are the third factor that ensures a correlation between moral truths and our moral beliefs.

Unlike Copp, however, Wielenberg further provides a more specific reason for why the EDA can thus be rejected: The third factor explanations prevent the EDA from providing a defeater for the epistemic status of our moral beliefs. In Chapter 4, I already

[135] Some people may argue that we would still have the right to life even if we did not have the relevant cognitive faculties. For example, we commonly agree that children and some animals also have the right to life. However, this kind of right to life would not be the same kind of a moral right as the one discussed by Wielenberg that is grounded in one's cognitive faculties. See also Griffin (2008: 83-95).

explained why the fact that our moral beliefs are not tracking the moral truth due to their evolutionary origin can provide an undercutting defeater, which is also a higher-order defeater at the same time, for our moral beliefs. In the same chapter, I also concluded that, according to the EDA, the epistemic status of our moral beliefs is undermined due to that undercutting defeater. Nevertheless, Wielenberg argues that the EDA fails to provide a defeater for the epistemic status of our moral beliefs because of the third factor explanations. Precisely speaking, he thinks that, if his third factor explanation is correct, then there is an indirect yet reliable correlation between moral truths and moral beliefs that cannot be undermined. If this is right, there is no reason for us to doubt the justification for our moral beliefs (Wielenberg 2014: 160).[136]

As an illustration of how Wielenberg's argument against the EDA works, we can start from why evolutionary debunkers think that there is an undercutting defeater for our moral beliefs. Let us assume that Mario used to believe that France won the World Cup in 2018. We can also imagine that there was a pill that could make him have that belief. Let us also assume that Mario was quite certain that he had taken this belief pill, but he then suddenly discovers the effects of this pill. As the evolutionary debunker Richard Joyce (2006) suggests, Mario's discovery of the effects of this pill should undermine his confidence in his previous belief (179). More precisely, the fact that he ingested the belief pill gives him a reason to doubt the reasons and evidence that led him to believe that France won the World Cup in 2018 in the first place.

[136] Interestingly, Enoch (2011) mentions in a footnote that the Third Factor Objection may provide a defeater-defeater which can defeat the defeater for our moral beliefs suggested by the evolutionary debunkers (170n41). Unfortunately, he does not further explain how the Third Factor Objection could provide such a defeater-defeater.

If this is right, then it seems to make sense that the justification Mario had for believing that France won the World Cup in 2018 is undermined by an undercutting defeater in this case. Following the same line of reasoning, it could be suggested that the connection between our moral beliefs and the evolutionary forces is the same as the connection between Mario's belief that France won the World Cup in 2018 and the belief pill in the previous case (Joyce 2006: 181).[137] That is, it can be argued that the fact that the evolutionary forces have influenced our moral beliefs gives us a reason to doubt whatever justifications we seemed to have for our moral beliefs and thereby this fact seems to function as an undercutting defeater for the epistemic status of those beliefs.

We can then return to why Wielenberg suggests that the EDA fails to provide such an undercutting defeater for our moral beliefs because of the third factor. Let us imagine that Mario actually took a belief pill called the Third Factor Pill. When one ingests this pill, there are two "salient causal consequences" (Wielenberg 2014: 160). Firstly, in that person's mind, the pill will produce a belief that France won the World Cup in 2018. Secondly, the pill will also cause the France national football team to defeat all opponents and to win the World Cup in 2018 "via a process of backward causation" (Wielenberg 2014: 160).

In this situation, as Wielenberg notes, there is no causal or direct correlation between the fact that France won the World Cup in 2018 and Mario's belief that France won the World Cup in 2018. However, his previous belief and the fact in question are correlated

[137] I also discussed the example of belief pill in Chapter 1 and sub-section 4.2.3.

because they are both "effects of a common cause – the ingesting of the [Third Factor Pill]" (Wielenberg 2014: 160). Therefore, even if Mario discovered that he ingested the Third Factor Pill, this fact in this situation would not generate a defeater because there is no reason to doubt the justification for the previous belief that he had in the first place.

As mentioned previously, Wielenberg's version of the Third Factor Objection likewise suggests that the third factor is responsible for both moral truth and moral beliefs. It can therefore be argued that, even if we discovered that the evolutionary forces have influenced our moral beliefs, this fact fails to generate any undercutting defeaters for our moral beliefs (Wielenberg 2014: 161). This is because that fact alone would not be able to provide any reason to doubt the correlation between the moral beliefs and moral truth because this correlation is already explained by the presence of the third factor. As long as our moral beliefs about moral rights are produced by the "right sort of process" – that is, by the relevant cognitive faculties, it is unlikely that their epistemic status would be undermined (Wielenberg 2014: 148-149).

6.5 David Enoch's Goodness of Survival

Finally, we can turn to how David Enoch answers the questions TF1-4 and thus his version of the Third Factor Objection.

6.5.1 Enoch's Reply to Question TF1

Enoch concedes that the evolutionary forces have influenced our moral beliefs. However, he suggests that evolution "aims" at the survival of a species when it influences the species. When he uses the term "evolutionary aim", he takes it "as [a]

shorthand for the usual respectable, non-teleological, evolutionary way of putting things" (Enoch 2011: 168n38). However, this definition is unclear. After all, what could this non-teleological, evolutionary way of putting things be? Unfortunately, Enoch does not explain it further. By using the term 'evolutionary aim', I believe that Enoch wants to emphasize that evolution does have a kind of guiding force towards a certain direction (James 2011: 126-127). We have evolved to have certain traits because there is a causal influence from the evolution that pushes us in the direction of having certain traits. I think that this kind of causal influence is what Enoch regards as a kind of non-teleological, evolutionary way of putting things, or, in other words, as the evolutionary aim.

Let us consider giraffes and their long necks as a non-moral example. According to Enoch's view, giraffes have evolved to have long necks because evolution influences them with the aim of survival. The long necks have been a fitness-enhancing trait for the giraffes in their natural environment. Evolution has then pushed giraffes to have long necks because they enable giraffes to reach the top of tall trees that other species could not possibly reach (Darwin 1876/2009: 177).[138] As a result, evolution influences giraffes to have this fitness-enhancing trait that enables them to access more food, and this is because it influences them with the aim of survival of the species. On this view of evolutionary aim, organisms have evolved to have certain traits that are fitness-enhancing because evolution pushes them to have these traits.

[138] Robert E. Simmons and Lue Scheepers (1996) doubt the view that giraffes have evolved to have long necks because this trait can help them to access more food. Instead, they argue that long necks may help male giraffes to better "fight for dominance and access to female [giraffes]" (Simmons & Scheepers 1996: 771).

We can then return to the discussion of the third factor. Enoch speculates that it could be thought that "survival or reproductive success is good" itself is the relevant third factor (Enoch 2011: 168). This is to say that he assumes that survival and reproductive success have certain positive, *pro-tanto* value (Enoch 2011:168). If Enoch's speculation is right, then it is also a normative fact that the evolutionary forces "aim" at something that is good (Enoch 2011: 168). Hence, Enoch suggests that the normative fact – the goodness of survival or reproductive success – is the third factor that is responsible for the correspondence between our moral beliefs and moral truth (Wielenberg 2010: 450; Enoch 2011: 168).[139]

6.5.2 Enoch's Reply to Question TF2

Enoch also makes an evolutionary claim about the connection between our moral beliefs and the third factor. He suggests that the evolutionary forces have indeed pushed us to have evolutionarily beneficial moral beliefs (and normative beliefs more generally) with the aim of reproductive success (Enoch 2011: 169). Let us consider the moral belief that wanton killing is wrong as an example again. In the previous 2.4.2 and 3.3.3, I mentioned that most of us, if not all, believe wanton killing is morally wrong. On Enoch's view, we have come to have this moral belief because the evolution aims at the survival and reproductive success of our species too. Because of this moral belief, the act of wanton killing has been generally regarded to be impermissible in all societies. In other words, members of almost all societies also have this moral belief. It is then

[139] Enoch (2011) actually argues that the goodness of survival is a third factor connecting the normative beliefs and normative truths (168). For the sake of discussion, in this thesis, I will merely focus on the cases of moral beliefs and truths instead of more general normative beliefs and truths.

reasonable to think that the prohibition of the act of wanton killing has increased the survival rate of human beings, which could be understood as the evolutionary aim.

As I mentioned in 6.5.1, Enoch also relies on the normative thesis that survival is good. Given this normative thesis and the previous evolutionary claim, the evolutionary forces can be thought to have pushed us to have evolutionarily beneficial moral beliefs that are also *pro-tanto* good to have (Enoch 2011:169). Enoch (2001) uses the moral belief that pain is *pro-tanto* bad as an example (169). According to Enoch's evolutionary claim, the evolutionary forces pushed human beings to have this belief because having it enabled us to avoid the kind of injuries that often lead to death. Hence, it is also *pro-tanto* good for us to act in the way that helps us to avoid pain as this will also help us to survive, which is a good thing, according to Enoch's normative view.

If Enoch is right, then there is a correlation between the goodness of survival (i.e., the third factor) and the content of our moral beliefs. This is because when the evolutionary forces influenced our moral beliefs, they influenced them in a way that promotes what it is *pro-tanto* good according to Enoch's normative assumption (2011: 169).

6.5.3 Enoch's Reply to Question TF3

In order to explain why the third factor is correlated with the moral truth, Enoch makes another normative claim. That is, he suggests that there are coherence relations between the normative fact that survival and reproductive success is good and many other normative truths (Enoch 2011: 168-70). Let us consider the proposition that "pain is *pro-tanto* bad", which is an example of a normative truth according to Enoch (2011:

169). He argues that there is a coherence relation between the normative fact that pain is *pro-tanto* bad and the normative fact that survival and reproductive success is good (i.e., the third factor) (Enoch 2011: 169). After all, avoiding a bad thing (pain) here also promotes a good thing (survival) at the same time in this case.

However, it should be noted that Enoch does not actually provide a clear definition of what coherence really consists in this context. I will come back to discuss this question in the sub-section 6.6.4 below. For the sake of simplicity, I will provide a very basic definition of the coherence relation here. According to this preliminary definition, a fact X coheres with a fact Y, if and only if, (i) X is logically consistent with Y, and (ii) there is a relation of mutual support between X and Y. Firstly, the fact that pain is *pro-tanto* bad and the fact that survival is good are not logically inconsistent. Secondly, we tend to avoid dangers because pain is *pro-tanto* bad, and avoiding dangers is certainly beneficial for the survival of our species. Hence, it seems that there is a relation of mutual support between these two facts too. As a consequence, arguably there is a coherence relation between the normative fact that survival is good and the third factor according to the preliminary definition of coherence. Enoch further insists that many other normative truths also cohere with the normative fact that survival and reproductive success is good, perhaps in the way described above.

6.5.4 Enoch's Reply to Question TF4

According to the EDA, if our moral beliefs do not adequately align with the corresponding moral truths (i.e., track the moral truth reliably) due to their evolutionary origin, then the epistemic status of these beliefs is undermined. Enoch himself also

agrees that, if moral realists fail to provide a plausible explanation of the alignment between our moral beliefs and the moral truth, then it would be difficult for them to think that our moral beliefs are actually formed in a reliable way (Enoch 2011: 158-160). Hence, he also agrees that the epistemic status of our moral beliefs is undermined if the moral realists fail to explain the alignment in question. However, he then argues that such an alignment can actually be based on the third factor and, as a result, the epistemic status of our beliefs is not undermined.

Based on his answers to TF1-3, Enoch argues that our moral beliefs, which are formed under evolutionary influence, are "systematically related" and "somewhat in line with" the moral truths (Enoch 2011: 168-69), and this is because the fact that survival is good – the third factor – "pre-establishes the harmony between the normative truths and our normative beliefs" (Enoch 2011: 168). Again, Enoch does not further explain what he really means by a systematic relation or the pre-established harmony. However, I suggest that the systematic relation can be considered to be a kind of an indirect relation between the moral truth and our moral beliefs. I also suppose that, when the third factor is responsible for both the moral truth and the moral beliefs, this third factor can be thought to be pre-establishing the harmony between them.

According to Enoch, the indirect correlation (or harmony) between moral beliefs and the moral truth can be considered to be an adequate alignment that is reliable enough to avoid the sceptical worries of the EDA. Just like Wielenberg, on Enoch's view, even if the evolutionary forces had significantly influenced these beliefs, this fact fails to generate any undercutting defeaters for our moral beliefs. This is because this

evolutionary fact would not be able to provide any reason to doubt the correlation between the moral beliefs and moral truth because this correlation is already explained by the indirect correlation pre-established by the third factor. Therefore, the epistemic status of our moral beliefs would not be undermined even if what I argued in Chapter 4 were right.

6.6 Reply: A Trilemma for the Defenders of the Third Factor Objection

The aim of this Chapter 6 is to defend the EDA against the Third Factor Objection. In this §6.6, I will argue that the third factor explanations suggested by Copp, Enoch and Wielenberg are all problematic for different reasons. Therefore, at least so far we have not been provided with a version of the Third Factor Objection that would be a satisfying response to the EDA. §6.6 will outline how the defenders of the Third Factor Objection face a trilemma and, as a consequence of this trilemma, they face a genuine difficulty in being able to formulate an acceptable version of the Third Factor Objection. The basic crux of my argument against the Third Factor Objection, which has the form of a trilemma, can be formulated as follows:

An Argument against the Third Factor Objection

1. The Third Factor Objection is a response to the EDA, which is predominantly used by the moral realists to defend their metaethical position.

2. There are three predominant versions of the Third Factor Objection, which

have been formulated by Copp, Enoch and Wielenberg respectively.

3. However, we cannot accept these versions for the following reasons:

3.1 If moral realists accept Copp's version of the Third Factor Objection, they need to give up their commitment to realism – the view that moral truths are attitude-independent.

3.2 If we accept Wielenberg's version of the Third Factor Objection, we will need to posit a brute and inexplicable relation between the third factor and moral truth which is too implausible for us to accept.

3.3 If we accept Enoch's version of the Third Factor Objection, we need to concede that many moral beliefs that are obviously false can too be justified.

4. An acceptable version of the Third Factor Objection needs to be able to avoid these three problems.

Conclusion: It would be difficult to formulate an acceptable version of the Third Factor Objection (1, 3, 4).

Let us start from considering the premise 1. There are two reasons to believe that the Third Factor Objection is predominantly adopted by the moral realists. First of all, *all*

the prominent defenders of the Third Factor Objection – including Copp, Wielenberg and Enoch – are moral realists, or at least they regard themselves as moral realists. When they defend the Third Factor Objection, they all openly admit that they hold a realist meta-ethical theory. Firstly, Copp (2008) claims that the society-centered theory, which is the meta-ethical theory that he defends, can be considered to be "a realist theory" which holds that there are moral facts. (203). As for Wielenberg (2014), he explicitly suggests that he holds the meta-ethical view of "robust normative realism", which is obviously a version of normative realism (including moral realism) (13). Finally, Enoch (2011) too suggests that he holds the meta-normative view that can be characterized as "robust realism" (1).

Secondly, in the previous 3.2.1 and 5.4.2, I already explained that the EDA is a targeted argument and moral realism is widely accepted to be the most significant target of the EDA (Street 2006: 110). As a response to the EDA, the Third Factor Objection can thus be thought to be an objection that is predominantly adopted by the moral realists as a way of defending their metaethical position from the EDA challenge. Indeed, Copp, Wielenberg and Enoch all try to offer the third factor explanations in order to defend moral realism from the EDA. When Copp (2008) introduces his response, he suggests that he has argued that "realists who accept the Darwinian hypothesis do not need to accept the tracking thesis to avoid a skeptical result" (194). Also, Wielenberg (2014) likewise suggests the EDA provides a difficult challenge to realists, and his version of the Third Factor Objection can help realists to deal with that challenge (155-156). Lastly, Enoch (2011) also suggests that his version of Third Factor Objection is a "realist way of coping with the epistemological challenge" including the EDA (165).

As a result, we should understand the Third Factor Objection as an objection which the moral realists use to deal with the EDA in order to defend their realism. The premise 1 of the argument against the Third Factor Objection can thus be thought to be true. Besides, I already explained why the premise 2 of the argument under investigation can be considered to be true in the previous §6.3-5. In the sub-sections 6.6.1-3 below, I will explain the premise 3 in detail – I will argue why Enoch's, Copp's and Wielenberg's versions of the Third Factor Objection all fail for different reasons. Finally, I will explain the premise 4 and conclude why my argument against the Third Factor Objection really work in the conclusion §6.7.

6.6.1 First Horn of the Trilemma: Why Copp's version of the Third Factor Objection Fails

Although David Copp regards himself as a moral realist, it is a concern that most moral realists cannot reasonably accept his version of the Third Factor Objection because of their commitment to moral realism. To explain this worry, let us start from just how Copp defines moral realism. According to Copp (2001), moral realism consists of two core theses – (i) our moral claims express moral propositions that are either true or false and (ii) some of the moral propositions are really true (223). This definition of moral realism suggested by Copp is a very broad one, and most moral realists would too accept the previous two theses. Nevertheless, the problem with this definition is that it is over-inclusive. It appears to include under moral realism the meta-ethical views that make moral truths attitude-dependent. This means that, according to this definition of moral realism, even traditional forms of meta-ethical subjectivism and relativism

should be considered to be versions of moral realism.[140]

Let us consider subjectivism as an example. In the previous 5.4.2, I already explained that subjectivists hold the view that moral utterances are reports of the speakers' attitudes and the truth of those utterances depends on their attitudes (Köhler 2012: 73; Miller 2003: 37). For instance, when I claim that 'kindness is good', my moral utterance is true if it accurately reflects my positive attitude towards kindness. According to subjectivism, moral utterances are literally true or false and some of them are true as long as they accurately report the speakers' attitudes. This means that Copp's definition of realism would have to include subjectivism to be a moral realist view.

Yet, the majority of moral realists would not consider subjectivism as a form of moral realism. As mentioned in the previous 3.2.1 and 5.4.2, most moral realists hold the essential view that moral truths are attitude-dependent. In other words, on their view, whether a moral utterance is true or false is independent of the speaker' attitudes (Finlay 2007: 820-822). For example, according to most realists, my moral claim that kindness is good can still be true even if I do not have any positive attitude towards kindness. As a result, if Copp's definition of moral realism includes subjectivism and any other meta-ethical views that make moral truths attitude-dependent, then it can be argued to be over-inclusive.

[140] Interestingly, in a recent article "Moral relativism is moral realism", Gilbert Harman (2015) argues that his version of meta-ethical relativism, which is "a claim about reality", can too be considered to be a version of moral realism (858). However, I do not think that most moral realists would agree with Harman as his definition of moral realism is over-inclusive.

In §6.3, I explained why Copp thinks that moral truths are correlated with the third factor – because they are dependent upon the third factor, that is, the justified social moral code. Nevertheless, if the moral truths are really correlated with the third factor in the way described in §6.3, then it can also be argued that moral truths will, according to his view, be attitude-dependent, or so I will argue in the rest of this sub-section 6.6.1. This means that, if anyone wants to adopt Copp's version of Third Factor Objection, she also needs to accept the thesis that moral truth is attitude-dependent. However, most moral realists cannot really accept this thesis precisely because of their commitment to moral realism. Hence, moral realism is incompatible with Copp's version of the Third Factor Objection, unless moral realism is defined in the over-inclusive way that Copp defines it. As a consequence, I believe that Copp's objection fails to help the moral realists to give a plausible response to the EDA.

The crucial question for my response to Copp's objection then is: Why do we need to accept that moral truth will turn out to be attitude-dependent if we adopt Copp's version of Third Factor Objection? In 6.3.3, I explained that, according to Copp, there is a grounding relation between the justified social moral code (i.e., the third factor) and moral truth. Copp, after all, argues that a moral proposition is true only if it corresponds to a moral standard that is a part of a justified social moral code (Copp 2001: 219-223). In other words, moral facts are determined by the social moral code that can be justified for the members of the relevant society given their basic needs and values.

As I mentioned in 6.3.1, on Copp's view, a social moral code is justified if and only if the members of the society would have rationally selected this code (Copp 2001: 165).

At this point, we need to have a closer look at how Copp understands the notion of rationality. As he suggests, "rationality is a matter of being responsive to reasons" (Copp 2011: 176). That is, when we attempt to rationally choose a moral code, we need to consider what reasons we have for choosing it. Moreover, Copp relies on the "needs-and-values theory" of reasons to explain the underlying reasons that we have for choosing a certain moral code. According to this theory, needs and values provide self-grounded reasons for the members of a society to rationally select a moral code (Copp 2001: 168). Simply put, self-grounded reasons are the reasons that must be considered in order for one to live a minimally rational life. Hence, according to the needs-and-values theory, needs and values ground the reasons that you must consider when you rationally choose a moral code.

However, I think that accepting needs-and-value theory eventually makes Copp's version of the Third Factor objection problematic. The key question is why the members of different societies would be able to rationally select different social moral codes. We can start from the basic needs. Basic needs generally seem to be the same for the members of all societies. Consider social moral codes that consist of a moral standard forbidding harming others as an example. Although different cultures may have different ideas of what harm is, it is likely that everyone would rationally want to live a life that does not include being harmed. For example, we would rationally want to live lives in which our arms would not be broken by others. And the need for not to be harmed provides a self-grounded reason for every member of all societies to select social moral codes that consist of a moral standard that forbids the acts of harming others. Therefore, we can expect that the members of all societies would rationally

select a social moral code that includes the moral standard forbidding the acts of harming others for that self-grounded reason. Following the same line of reasoning, it is also plausible to think that different societies would rationally select social moral codes that serve the same common basic needs.

Nevertheless, it is insufficient for a social moral code to be justified if it merely serves the basic needs of the members of the society. Let us consider the basic need not to be killed as an example. One can imagine that there will be many substantially different social moral codes that can satisfy our basic need not to be killed equally. We may then ask: Which one of these social codes could the members of a given society rationally choose? If they choose a particular social moral code merely based on their desires, they might not end up with a stable code. Thus, Copp suggests that they would rationally choose the social moral code that would satisfy not only their basic needs and but also their values given that values are more stable than mere desires. This is why Copp thinks that values too can provide self-grounded reasons for the members of a society to choose a social moral code rationally. For example, if the members of a society value the virtue of friendship, then they would have a good reason to choose a social moral code that forbids the act of manipulating our friends for our self-interest. In other words, according to the needs-and-value theory, rationality also seems to require us to choose social moral codes that would satisfy our values, as well as our basic needs.

However, it is unlikely that members of different societies would rationally select the same social moral codes because it is difficult to accept that the same values are shared

between the members of all societies (in the same way that needs are shared). To make things worse, values can conflict, and different societies may thus rationally subscribe to conflicting social moral codes in virtue of the conflicting values of their members. Indeed, different people often hold different and conflicting values even after rational reflection. For example, in normative ethics, many philosophers value pleasure and welfare after rational reflection, but there are also many philosophers do not particularly value pleasure and welfare, for example, Kantian ethicists and divine command theorists.

Let us further consider why members of different societies would rationally select different social moral codes in virtue of the values that they have. Let's take the virtue of filial piety – the respect for one's elders – as an example. It is true that most societies value filial piety to a certain extent. However, members of Chinese society value this virtue of filial piety to a great extent. We also call this Chinese virtue of filial piety Xiao 孝. When Edward Slingerland translates Confucius's *Analects* ([2003]), he emphases that Xiao is not just a virtue of "being a dutiful and respectful son or daughter", but this virtue also means that there is a debt that the son or daughter owe to their parents (238).

We can now consider the traditional Chinese social moral code that consists of moral standards that require sons and daughters to pay significant tribute to their parents. Firstly, in the ancient Chinese, when a parent died, the sons and daughters were required to mourn the death of their parents for three years. During this mourning period, they were not allowed to do a lot of things – including visiting friends, eating meat, drinking alcohol, attending public exams or working for the government. The underlying reason

for this mourning period was that the members of the traditional Chinese society, under the influence of Confucian philosophy, valued the virtue of Xiao.[141] Even today, the members of the Chinese society are still required to visit their ancestors' tomb at least once per year during the Ching Ming festival *清明節* because they still highly value the virtue of Xiao.

Another famous example of the traditional Chinese society valuing Xiao can be found in the Confucius's *Analects*. There was a son who reported his father to the authorities because the father stole a sheep, and Confucius did not find this son "upright" (Confucius [2003]:147). Instead, Confucius argues that an upright son should cover up the crime of his father even if the father really did something wrong (Confucius [2003]:147). Hence, members of a society highly valuing the virtue of Xiao would rationally select a social moral code that consists of a moral standard that requires sons and daughters even to cover up for their parents' criminal acts.

We can now return to the key question of why the members of different societies would rationally select different social moral codes with the example of the virtue Xiao. Let us imagine that there are a society A that highly values Xiao and another society B that does not value Xiao. Let us too assume that basic needs are the same in these two societies. To recap, according to Copp's needs-and-value theory (2001), a social moral code is thought to be justified for a society if and only if the members of this society would rationally subscribe to this code on the basis of what their needs and values are

[141] According to Confucius ([2003]), "a child is completely dependent upon the care of his parents for the first three years of his life – this is why the three-year mourning period is the common practice throughout the world" *子生三年，然後免於父母之懷。夫三年之喪，天下之通喪也* (210).

(2001: 101). However, given that members of the societies A and B have different values, the social moral codes that are justified to them would thus have different elements.

Let us then consider the moral principles that sons and daughters should go through a long mourning period for their dead parents and they should also cover up their parents' criminal acts. The members of the society A could rationally select a social moral code that included both of these principles because this society highly values Xiao and so the moral code that included the previous principles would best satisfy the needs and values of the society A. In contrast, the members of the society B could not rationally select the same social moral code because they do not value Xiao. As a result, this social moral code can be justified to the society A but not to the society B. This means that, on Copp's view, whether a social moral code can be justified for a given society is at least partly dependent on the values that the members of this society have.

Let us then consider the moral propositions 'it is right for sons and daughters to go through a long mourning period for dead parents' and 'it is right for sons and daughters to cover up the criminal acts of their parents'. To recall, Copp suggests that these moral propositions are true only if and just because the relevant acts are encouraged by the standards that are included in the justified social moral code. However, according to the same view, these two moral propositions are both true in relation to the society A but false in relation to the society B. This is because, in this example, the society A would have a justified social moral code including the standard that encourages the previous kinds of acts but Society B would not. Thus, some moral facts will be different in the

society A and society B depending on the different attitudes the members of these two societies. However, this also means that according to Copp moral facts are attitude-dependent.

As a result, if moral beliefs were correlated with the moral truths in the way described by Copp's third factor explanation, then most moral realists will be unable to accept his version of the Third Factor Objection to the EDA. This is because they cannot accept that moral truth would be attitude-dependent in the way entailed by Copp's view. The cost of accepting so would, after all, be giving up moral realism.

On the basis of this problem, we can draw a more general lesson. We can ask: What does an acceptable version of Third Factor Objection have to be like? The previous discussion began from the observation that moral realists are fundamentally committed to the claim that moral truths are attitude-independent. Because of this observation, the general lesson that we can draw from Copp's version of the Third Factor Objection is the following: If a given third factor determines the moral truths in a third factor explanation, then human attitudes must not be a part of the third factor itself. Likewise, the connection between the third factor and the moral truths should also be attitude-independent. If the human attitudes were part of the third factor that determines the moral truths, or the connection between moral truth and the third-factor were attitude-dependent, then the moral truths too would become attitude-dependent, which is something that the moral realists cannot accept given their fundamental commitment to attitude-independent moral truths.

6.6.2 Second Horn of the Trilemma: Why Wielenberg's Version of the Third Factor Objection Fails

Let us then focus on Wielenberg's version of the Third Factor Objection. Unlike Copp, Wielenberg avoids the previous problem of Copp's view because the third factor and moral truths are attitude-independent according to his account.[142] His version of the Third Factor Objection, however, is problematic for a different reason. The relevant problem is that, according to Wielenberg, moral truths are connected to the third factor in a certain special way that is brute and inexplicable.

Let's start from just how Wielenberg thinks that the moral truths are correlated with the third factor. In 6.4.3, I mentioned that, according to his view, the presence of our moral rights (which are a kind of moral truths) is guaranteed by the presence of our cognitive faculties (which are the third factor). Furthermore, he understands this relation as an instance of what he calls the 'D-supervenience relation'. In other words, our moral rights are correlated with the relevant cognitive faculties because the former D-supervenes upon the latter (Wielenberg 2014: 145).

What then is the D-supervenience relation? It is a different kind of a relation compared to the ordinary supervenience relations. Wielenberg introduces the D-supervenience relation in order to deal with the supervenience challenge against moral realism. That is, he tries to explain ordinary supervenience of the moral on the non-moral by suggesting that the moral properties actually D-supervene upon non-moral properties.

[142] Firstly, Wielenberg (2014) explicitly accepts that there exist response-independent normative truths (13). As for the third factor, although he does not explicitly suggest that cognitive capacities are attitude-independent, it seems that our cognitive capacities can arguably be considered to be attitude-independent.

Before I explain the D-supervenience relation, let me first explain the ordinary supervenience relation in ethics. Moral supervenience is a logical relation according to which necessarily, when something has a moral property, it must also have some base property and, as a matter of metaphysical necessity, anything that has the same base property must also have the same moral property as well (Väyrynen 2018: 174). To illustrate the supervenience relation, let's consider Figure 6.2 below:

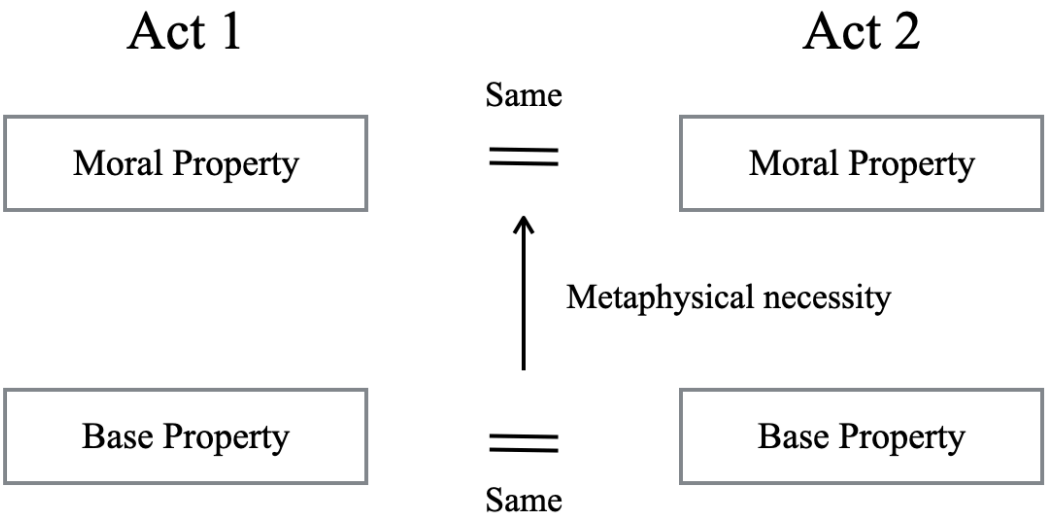

Figure 6.2 – Moral Supervenience Relation

Let us take the act of killing an innocent baby as an example. Let us call this act ACT_1 and all the base properties of the ACT_1 together as the property P. We can then consider another act of killing an innocent baby that we call ACT_2, and assume that this ACT_2 has all the same base properties P that the ACT_1 has and only them. Then, the supervenience relation entails that whatever moral properties the ACT_1 has, the ACT_2

also has the same moral properties.

Let me then explain the supervenience challenge against moral realism. This challenge, which is originally formulated by Simon Blackburn (1971: 101-124), is thought to be a difficult explanatory challenge for the non-naturalist realists (McPherson 2012). According to the non-natural realists, moral properties have an independent metaphysical status because they are *sui generis* and thus "discontinuous" with other kinds of properties, such as the natural and supernatural properties (Väyrynen 2018: 171; Wielenberg 2014: 14). As a consequence, moral properties should be considered to be ontologically distinct and separate from the base properties. Yet, this also means that it is very difficult for the non-natural realists to explain the basic supervenience relation between the moral properties and other properties.

Let's consider the example of the act of killing an innocent baby again. Under non-naturalist realism, it seems to be possible that the ACT_1 killing an innocent baby and the ACT_2 killing an innocent baby could have different moral properties, even if both acts had exactly the same base properties P. This is because if moral properties were metaphysically different and separate from the base properties, it seems to be logically possible that the two acts could have different moral properties even if they had exactly the same base properties P (Väyrynen 2018: 176). It is implausible to accept this consequence because it is hard to understand how the ACT_1 and the ACT_2 could have different moral properties, given that they share the same base properties. This means that the realists will need to be able to explain just why ACT_1 and the ACT_2 could not have different moral properties in this case and thus just why supervenience holds in

this case. As a result, the supervenience challenge is thought to be a difficult explanatory problem for the non-naturalist realists.

As a non-naturalist realist who wants to deal with the supervenience challenge by explaining the metaphysically necessary relation between the moral properties and the base properties, Wielenberg argues that the relation between moral properties and the base properties should be understood as a D-supervenience relation, which will then help us to explain the ordinary supervenience relation described above. Wielenberg (2014) takes the D-supervenience to be a special kind of making or causal relation, originally described by Michael R. DePaul (and thus the "D" of D-supervenience stands for "DePaul") (13, 17-18). According to Wielenberg (2014), when some moral properties M D-supervene on some base properties, M are dependent on the properties B in the way that "B's instantiation … makes M [to] be instantiated" (Wielenberg 2014: 11).

Once again, let us consider the act of killing an innocent baby as the example. In this example, the ACT_1 killing an innocent baby and ACT_2 killing an innocent baby both have the same base properties P. Let us further suppose that the ACT_1 has the moral properties M. On Wielenberg's view (2014), if the moral properties M D-supervene on the base properties P, this entails that the instantiation of P makes M to be instantiated (10-11). Let us assume that there really is a D-supervenience relation between moral properties M and base properties P and the ACT_1 has the moral properties M because it has the base properties P. In this case, the ACT_2 that has the same base properties P must also have the same moral properties M because the instantiation of P makes M to be

instantiated. As a result, Wielenberg argues that non-naturalist realists are able to explain the moral supervenience (i.e., why two acts could have the same moral properties given that they have the same base properties) by relying on the D-supervenience relation in the way described above.

We can now return to Wielenberg's version of the Third Factor Objection. To recap, according to his version of the third factor explanations, the presence of our cognitive faculties guarantees the presence of the moral rights. Furthermore, Wielenberg (2014) argues that this guaranteeing relation between our certain cognitive faculties and the moral rights can be characterized as an instance of the D-supervenience relation (155). In other words, he suggests that our moral rights are dependent upon our cognitive faculties because the presence of cognitive faculties makes or causes us to have those moral rights.

As mentioned 6.4.3, Wielenberg suggests that our advanced cognitive capacities enable us to perform activities and attain achievements that are valuable and unique to human beings. Since the possession of these cognitive faculties makes human beings so unique and valuable, as a result, it also makes us worth protecting and we thus have certain moral rights, such as the rights not to be killed and tortured. If Wielenberg (2014) were right and there really is a making relation (i.e., the D-supervenience relation) between the (i.e., the third factor) and moral truth (i.e., the moral rights), then it would be hard to deny that moral truth and the third factor are correlated in a very strong and robust way (155). This is because the making relation can also be understood as a special, robust sort of causation relation which can obviously entail a very strong kind of

correlation (Wielenberg 2014: 1).

However, I believe that Wielenberg's explanation of the relation between the third factor and the moral truth based on D-supervenience is unconvincing. This is because even if D-supervenience can arguably be responsible for a very strong kind of a correlation between the relevant cognitive faculties and the moral rights, that relation itself will be objectionably brute and inexplicable, or so I will argue next.

Let us consider why the D-supervenience relation is a strong relation that itself would require a further explanation. We can start from a comparison between ordinary supervenience and D-supervenience relation. In order to explain the ordinary supervenience relation between some moral properties M and base properties B, we are required to explain why M co-varies with B across possible worlds. In contrast, if some moral properties M D-supervene on some base properties B, this means that B actually has a power that can make M to be instantiated in all worlds (Wielenberg 2014: 20). Hence, in order to explain the D-supervenience relation between M and B, we do not need to explain why M and B co-vary, but rather we would need to be able to explain just how B has the power to make M to be instantiated necessarily (Wielenberg 2014: 20). If we fail to provide such an explanation, the D-supervenience relation should thus be considered to be a brute and inexplicable modal connection between distinct existences.

In fact, Wielenberg himself does not disagree that the D-supervenience relation is a brute connection, but rather he claims that it is not a problem that this relation is a brute

one. In his book *Robust Ethics*, Wielenberg (2014) himself concedes that moral realists need to commit to some brute facts that are unexplained because, in some cases, "explanation … must come to an end somewhere … Eventually we hit bottom; no further explanation is available" (24). Hence, he argues that, in some situations, it is not a problem for non-naturalist moral realists to be committed to some brute facts.[143] For example, he suggests that the D-supervenience relation between moral properties and base properties can be considered to be one of those brute facts that non-naturalist moral realists can happily posit (Wielenberg 2014: 24, 37).[144] Following the same line of reasoning, it is very likely that Wielenberg would also think that it is not a problem to posit a brute and inexplicable D-supervenience relation between our cognitive features and our moral rights.

However, I believe that it is a genuine problem that Wielenberg thinks that the connection between the third factor and the moral truth is a brute one because we should accept a methodological principle often called the "Modest Humean Commitment" (Väyrynen 2018: 176; McPherson 2012: 217). According to this principle, if a philosophical view claims that a given necessary connection (such as the supervenience relation) is brute, then the commitment to such a brute connection counts "significantly against its plausibility" (McPherson 2012: 217). This does not mean that all such

[143] According to Wielenberg, there are two reasons for non-naturalist realists to be committed to brute facts. Firstly, he thinks that even if a fact is a brute one, it does not necessarily mean that it cannot be proven to obtain (Wielenberg 2014: 37). Secondly, and more importantly, some brute facts can help the non-naturalist realists to deal with many explanatory challenges, such as the 'source problem of moral truths" (Heathwood 2012) and also the 'supervenience challenge" (Schroeder 2005: 3; Wielenberg 2014: 37).

[144] Wielenberg (2014) does suggest that the D-supervenience relation can be understood to be the "causal relation that many theists take to hold between a state of affairs being divinely willed and the obtaining of that state of affairs" (18-19). Nevertheless, the causation relation between God and His creation is also brute and inexplicable. As a result, even if we understood the D-supervenience relation in such a way, this would not mean that the relation in question would become less inexplicable.

philosophical views must always be rejected because perhaps there are some brute modal connections. However, all such philosophical views face a significant explanatory challenge: either they have to be able to provide an explanation of the necessary connection in question, or they have to show that their view has such great theoretical virtue that it is worthwhile to commit to the bruteness of the connection in question. For example, they could do the latter by showing that other alternatives are simply unattainable.

Let us take astrology as an example. Many people would want to be able to predict their destiny. In order to be able to do so, those who believe in astrology think that that there is a necessary relation between the zodiac signs and their destinies. However, such a view of a necessary relation is too brute and inexplicable to be acceptable. If a necessary relation is brute and inexplicable, then it seems that it is more reasonable to reconsider "one's commitment to the correlation" or even reject it, rather than accepting the bruteness of this correlation (Enoch & McPherson 2017: 834-835).

Following the same line of reasoning, the methodological principle "Modest Humean Commitment" can also be applied to Wielenberg's explanation of the relation between moral truth and the third factor in terms of D-supervenience. As mentioned above, Wielenberg takes the D-supervenience to be a metaphysical making relation that can be responsible for a very strong form of correlation between the relevant two types of properties. It is thus a very reasonable requirement that Wielenberg should be able to explain why there would be such a strong and robust relation between moral truth and the third factor and how that relation actually works. It is plausible to think that not

having such an explanation is a significant theoretical cost for the moral realists who end up having to posit the existence of this kind of brute connections. If the D-supervenience relation between the third factor and moral truth remains inexplicable, this should lower our credence in whether there is such a relation between the third factor and the moral truth in the first place (Enoch & McPherson 2017: 835).

Since I already explained why the D-supervenience relation between our cognitive faculties and our moral rights in Wielenberg's third factor explanations remains brute and inexplicable, our confidence in this explanation of the D-supervenience relation between the third factor and moral truth should also be decreased perhaps to the "point of suspension of judgment" (Enoch & McPherson 2017: 835). As mentioned in 6.2.3, all defenders of the Third Factor Objection are already in an explanatory debt to explain the correlation between the third factor and the moral truth. Wielenberg is merely shifting the explanatory burden from one place to another if he merely uses the D-supervenience relation to explain that correlation. More precisely, his solution is an attempt to solve the initial explanatory problem by "positing exactly the same sort of 'puzzling' … relation to do the explanatory work" (Enoch & McPherson 2017: 833). As a result, there is still a challenging explanatory challenge to anyone who wants to accept Wielenberg's version of the third factor explanation. It is therefore unreasonable, or at least there are no sufficient reasons for us to posit that there is a D-supervenience relation between moral truth and the third factor. In other words, it seems that Wielenberg's version of the Third Factor Objection is too implausible to be acceptable.[145]

[145] The D-supervenience relation, which is a type of a making or a causal relation, is similar to the

On the basis of this problem, we can draw another general lesson. The previous discussion began from the question of what the connection between the third factor and the moral truth could be like. On the basis of my objection to Wielenberg's view, the general lesson that we can draw is the following: The correlation between the third factor and moral truth cannot be brute and inexplicable. This is because if that correlation were a brute connection, then the defenders of the Third Factor Objection are just shifting the explanatory burden from one place to another and there would remain an explanatory challenge to them.

6.6.3 Third Horn of the Trilemma: Why Enoch's Version of the Third Factor Objection Fails

According to the two general lessons drawn from 6.6.1-2, there are two conditions that the defenders of the Third Factor Objection must meet when they put forward the objection. Firstly, the third factor and the moral truths need to be attitude-independent (6.6.1). Secondly, the correlation between the third factor and the moral truths cannot be a brute connection (6.6.2). Perhaps in order to avoid the previous problems, David Enoch's version of the third factor satisfies both conditions. Firstly, he understands both the third factor and the moral truths as attitude-independent normative truths.[146] Hence,

metaphysical causal relation as discussed by Alastair Wilson (2018). Hence, in this case, the relation between our certain cognitive faculties and the moral rights may also be characterized as an instance of metaphysical causation, and so we can use the best accounts of such causation, such as Wilson's, to make sense of the nature of that relation. Despite this, my objection to Wielenberg still stands: If the relation of metaphysical causation that obtains specifically between our cognitive faculties and the moral rights remains brute and inexplicable, there are still no sufficient reasons for us to posit that there is such a relation between them.

[146] Firstly, Enoch (2007) explicitly suggests that there exist response-independent normative truths (21). Furthermore, according to him, the third factor is the normative fact that "survival is good". Thus, the normative fact that survival is good can also be considered to be response-independent as well.

the first condition is clearly satisfied. Secondly, he explicitly suggests that the relation between moral truth and the third factor is not a necessary metaphysical making relation (Enoch 2011: 169). Instead, he suggests the correlation between the third factor and moral truths is an ordinary coherence relation and thus not a brute connection (Enoch 2011: 169).

To further illustrate the idea of coherence relations between the third factor and the moral truths, let's consider the following figure which illustrates the basic crux of Enoch's version of the third factor explanation:

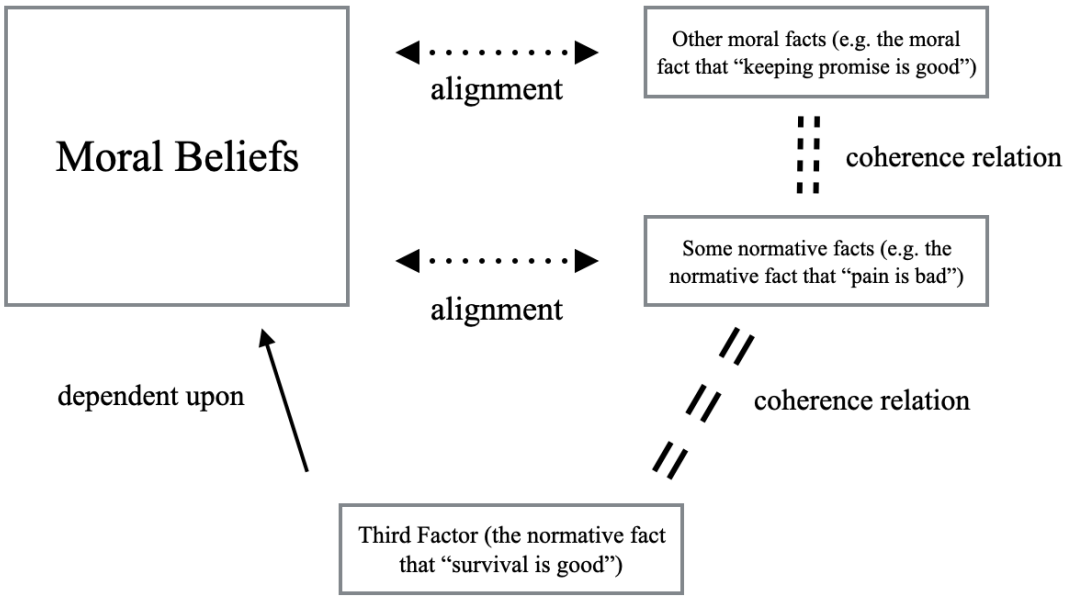

Figure 6.3 – Enoch's third factor explanation

As Figure 6.3 illustrates, on Enoch's view, normative facts correlate with the third factor due to a coherence relation (I will further explain what a coherence relation is below).

Moreover, the third factor itself is not directly connected to all moral facts but just coherently connected to some *specific* normative facts. This is because, according to Enoch, the third factor is the normative fact that survival is good, and this fact only coheres with certain *specific* normative facts, such as the fact that pain is bad and the fact that killing is bad. Many other moral facts (for example, the fact that keeping a promise is good) are then only correlated with the third factor because of a whole chain of indirect coherence relations via those *specific* normative facts.

Let us consider one of the *specific* normative facts, the fact that pain is bad. Believing that this fact obtains can be thought to an evolutionary advantage. For example, when the members of a given species hold this belief, they are able to avoid many dangers that would cause pain to them. Avoiding those dangers would enable more members of that species to survive, and their overall fitness would thus be enhanced. As a result, the normative fact that pain is bad coheres with the more basic normative fact that survival is good (i.e., the third factor). Moreover, Enoch (2011) suggests that this *specific* normative fact that pain is bad also coheres with many other moral facts – including the facts that causing physical harms to others is wrong and that torturing people is wrong (169). If Enoch is right, then many other moral facts too would also correlate with the third factor because of a whole chain of indirect coherence relations via certain *specific* normative facts such as the fact that pain is bad.

However, I believe that Enoch's third factor explanation is also unconvincing because of those coherence relations. Let us assume that Enoch is right and that survival is good really is a normative fact. The problem is that, even if the correlation between the

normative fact that survival is good (i.e., the third factor) and many other moral facts was based on merely a coherence relation, the resulting correlations would be insufficient for the purposes of the third factor explanation. Let us start from the nature of the coherence relation. Although there is no clear, universally accepted definition of what this kind of a relation is, there are two common accounts of what coherence could consist of (Audi 2011: 221):

- Account A: For X to cohere with Y, X is logically consistent with Y.
- Account B: For X to cohere with Y, (i) X is logically consistent with Y and (ii) there is a relation of mutual support between X and Y.

Let us first consider the coherence account A. On this view, for X to cohere with Y, X cannot be logically inconsistent with Y and they thus must both be able to be true at the same time. For example, the proposition '3+1=4' is inconsistent with another proposition '3+1=5' because two propositions cannot be both true at the same time. As a result, these two propositions cannot be in a coherence relation according to the account A.

However, if the third factor merely cohered with the moral truth where this means that the two merely are not logically inconsistent with one another, the third factor explanation would be too implausible to accept. Firstly, the third factor (i.e., the fact that survival is good) seems to cohere with both the proposition that 'pain is good' and proposition that 'pain is bad' at the same if we adopt the coherence account A. This is because neither one of these propositions are logically inconsistent with the third factor

fact that survival is good. Furthermore, the third factor would not merely cohere with certain moral facts. Rather, it would also cohere with many other moral falsehoods such as the propositions that killing is good and torture is good because none of them is logically inconsistent with the third factor fact. Hence, almost no moral beliefs that we have (no matter how absurd they were) would fail to be justified under the coherence account A. As a result, it is too implausible to understand Enoch's coherence relations according to the coherence account A.

We can then turn to the coherence account B. According to this account, for X to cohere with Y, firstly X cannot be logically inconsistent with Y. Yet, on this view, logical consistency is insufficient for X to cohere with Y. In addition, X must also explain Y in some way, or Y explain X, or X and Y must be mutually explanatory (Dancy 1985: 110-112). Let us consider the propositions that 'Peter is wet when he enters the building' and 'it is raining outside' as examples. According to the coherence account B, these two propositions are in a coherence relation. Firstly, they are not logically inconsistent. Secondly, and more importantly, it seems that the truth of the first proposition according to which Peter is wet when he enters the building supports the truth of the second proposition according to which it is raining outside. According to the account B, these two propositions are thus in a coherence relation as both conditions of logical consistency and mutual support are satisfied.

Nevertheless, even if Enoch understood the relevant coherence relations in terms of the account B, his version of the Third Factor Objection would still be implausible. Let's assume that the coherence relation is understood in accordance with the account B. We

can also grant that there really is a normative fact that survival is good and this fact coheres with at least some *specific* normative facts including the fact that pain is bad. However, even if we granted that the third factor could cohere with few such *specific* normative truths, the problem will be that the same will not be the case when it comes to many other important moral truths.

Let us consider moral proposition that 'it is right to kill human beings who lack the ability to contribute to the functioning of the society' as an example. The problem is that this moral proposition would cohere with the third factor fact that survival is good if we understood the coherence relation according to the coherence account B. In addition, if we all had the belief that it is right to kill people who lack the ability to work, it seems that the overall fitness of our species would be enhanced. It also seems that the third factor fact that survival is good could also explain why I would form that moral belief because having the previous belief could be thought to be an evolutionary advantage. Then, according to Enoch's version of the Third Factor Objection, the belief in question would too be justified if I happened to have it.

But this is an unwanted implication. The moral proposition that 'it is right to kill human beings who lack the ability to contribute to the functioning of the society' is obviously a moral falsehood. Following the same line of reasoning, it turns out that, under Enoch's version of the Third Factor Objection, the third factor fact that survival is good also cohere also with many moral falsehoods, including moral propositions like 'it is right to kill anencephalic infants' and 'it is right just to help your kin but not to help any strangers'. As a result, Enoch's third factor explanations would allow a lot of false moral

beliefs which at least some individuals and groups have formed under the evolutionary influence to be justified. This is why it seems that Enoch's version of the Third Factor Objection too is not sufficiently plausible.

On the basis of this problem, we can draw another general lesson. The general lesson that we can draw from Enoch's version of the objection is the following: The correlation between the third factor and the moral truths cannot be a coherence relation either.

6.7 Conclusion

In order to address the EDA, the Third Factor Objection argues that there are certain third factors that can guarantee that there is an indirect but yet reliable correlation between our moral beliefs and moral truth. If this were right, then the epistemic status of our moral beliefs would not be undermined due to their evolutionary origin. However, based on the three general lessons that we drew from the previous §6.6, there are three conditions that an acceptable version of the Third Factor Objection must meet:

1. If a given third factor determines the moral truths in a third factor explanation, then human attitudes must not be a part of the third factor itself or the relationship between the third factor and the moral truths. This is because otherwise the resulting moral truths would be attitude-dependent and so defending a third factor response of this type would require giving up moral realism.

2. The correlation between the third factor and moral truths cannot be brute and inexplicable.

3. The correlation between the third factor and the moral truths cannot be a coherence relation.

All things considered, this means that it is very difficult to formulate an acceptable version of the Third Factor Objection that can satisfy all three conditions above. After all, it is very hard to explain why and how moral truths could be plausibly correlated with the evolutionarily influenced third factor, whatever the third factor could be. One may attempt to argue that there is still a possibility that there could be such a version of the Third Factor Objection that could satisfy all three conditions. I do not rule out this possibility. Nevertheless, §6.6 already illustrates why it is a difficult task to provide a feasible version of the Third Factor Objection that could be reasonably acceptable. This means that until the defenders of moral realism are able to come up with such an objection there is no plausible third factor response to the EDA.

Chapter 7

Conclusion

In this thesis, I have focused on the epistemological evolutionary debunking arguments in meta-ethics (EDA). The main objective of my thesis has been to defend the EDA. Throughout this thesis, I have accomplished two main tasks to construct a defence of the EDA: (1) I have offered a more detailed and precise way to understand the EDA and (2) I have rejected the two strongest objections to the EDA. In this chapter, I will give a brief summary of how I have completed these two tasks in my thesis. In **Part I** of my thesis, I offered my account of how the EDA should be best formulated. First of all, in **Chapter 2**, I described how evolution has significantly influenced and shaped our moral beliefs as captured by the Adaptation Account and the Exaptation Account. I also concluded that the debunkers can wholly explain the evolutionary origin of those beliefs as long as one of the previous accounts above is true.

Then, in **Chapter 3**, I showed that our moral beliefs are not tracking the moral truth reliably given that the origin of those beliefs can be wholly explained in evolutionary terms. In that chapter, I started from asking of what truth-tracking actually consists. I then suggested that the liberal version of the explanatory reading of truth-trackingness and the modal reading of truth-trackingness can be used by the evolutionary debunkers to understand of what the required kind of alignment between moral beliefs and moral facts truth-trackingness could consist. In that chapter, I eventually concluded that most, if not all, of our moral beliefs can be argued to be not tracking the moral truth because

of their evolutionary origin as long as one of those accounts of truth-trackingness is true.

Finally, in **Chapter 4**, I addressed the question of why the epistemic status of our moral beliefs is undermined as a result of the fact that these beliefs are not tracking the moral truth. In that chapter, I argued that the fact that our moral beliefs are not tracking the moral truth reliably due to their evolutionary origin provides an undercutting and a higher-order defeater for those beliefs. As a result, I concluded that the epistemic status of our moral beliefs is undermined because of these two kinds of defeaters.

Based on what I argued in **Chapter 2-4**, I accomplished the first task of constructing the best possible formulation of the EDA. This formulation strongly supports the idea that the epistemic status of our moral beliefs is undermined due to their evolutionary origin.

In order to provide a comprehensive defence of the EDA, I focused on the second task in **Part II** of my thesis. In this part, I introduced and also argued against the two strongest objections to the EDA. Firstly, in **Chapter 5,** I focused on the Conceptual Truth Objection to the EDA. In that chapter, in response to this objection, I first provided a Hare-styled argument against the classical theory of moral concepts on which the objection relies and I also objected to the view that our moral beliefs could be justified by conceptually analyzing the moral concepts. Based on these two responses, I concluded that the Conceptual Truth Objection fails as an objection to the EDA.

Secondly, **Chapter 6** introduced the Third Factor Objection and also the three main versions of that objection by David Copp, David Enoch and Erik Wielenberg. In that chapter, I argued that these three versions of the Third Factor Objection are problematic for three different reasons. On the basis of these three problems, I argued that there are three conditions that a plausible version of the Third Factor Objection would have to be able to satisfy. I then suggested that, as a result, it would be very difficult to formulate a plausible version of the Third Factor Objection that can avoid those three problems. As a consequence, I concluded that there currently doesn't exist a plausible version of the Third Factor Objection, and it is also unlikely that such a version could be constructed in the future.

Based on the work done in this thesis, I concluded that the epistemic status of our moral beliefs is undermined due to their evolutionary origin. I also concluded that moral realists, particularly non-naturalist realists, thus face a serious epistemological challenge and I strongly believe that they are unlikely to be able to deal with that challenge in a plausible way. In line with this outcome, my future research project involves investigating the meta-ethical views according to which we could still have moral knowledge even if we did not have any justified true moral belief in the first place. I am particularly interested in evaluating quasi-realism (Blackburn 1984; 1988; 1993; 1996; 1998; 2006; 2009; Egan 2007; Ingram 2017) and revolutionary fictionalism (Joyce 2001: 175-231; 2005; Daly 2008) in the future.[147]

[147] More precisely, my future research project will focus on three specific issues regarding quasi-realism and revolutionary fictionalism. Firstly, I will investigate the recent developments in the Frege–Geach Problem literature, which is thought to be one of the strongest objections to the quasi-realism and also

In closing, philosophers and ethicists have a long tradition of working on interdisciplinary research. I strongly believe that they cannot completely turn a blind eye to the latest developments in other fields including different natural sciences, biology, psychology, sociology, and so on. From my point of view, the EDA is a very important meta-ethical project that can also be considered to be a proper response to the latest development in evolutionary biology. With this thesis, I have thus tried to make a contribution to an argument that I consider to be one of the most promising interdisciplinary research projects in meta-ethics in recent years.

ethical expressivism (Geach 1960; 1965; Blackburn 1984: 189-192; 1988: 504-508; 2006: 155-157; Sinclair 2009: 142-143). For some representative discussions of the Frege–Geach Problem in recent years, see Charlow (2014), Hung and Tse (forthcoming), Lennertz (forthcoming), and Woods (2017). Secondly, I will explore the question of whether quasi-realism should be considered to be a form of fictionalism (Blackburn 2005; Lewis 2005; Nolan et al. 2005). Finally, I will focus on reactionary fictionalism – which is recently suggested by Jason Dockstader (2020a; 2020b). Especially, I will attempt to assess whether this new version of fictionalism is more plausible than revolutionary fictionalism.

Bibliography

Alston, W. P. (1985). Concepts of Epistemic Justification. *The Monist, 68*(1), 57-89.

Anselm of Canterbury (1059/1998). Proslogion. In B. Davies & G. R. Evans (Eds.), *Anselm of Canterbury: The Major Works* (pp. 82-104). New York: Oxford University Press.

Artiga, M. (2015). Rescuing Tracking Theories of Morality. *Philosophical Studies, 172*(12), 3357-3374.

Audi, R. (2004). *The Good in the Right: A Theory of Intuition and Intrinsic Value*. Princeton, N.J.: Princeton University Press.

Audi, R. (2008). Intuition, Inference, and Rational Disagreement in Ethics. *Ethical Theory and Moral Practice, 11*(5), 475-492.

Audi, R. (2011). *Epistemology: A Contemporary Introduction to the Theory of Knowledge* (3 ed.). New York; Oxford: Routledge.

Audi, R. (2015). *Reasons, Rights, and Values*. Cambridge: Cambridge University Press.

Ayala, F. J. (2010). The Difference of Being Human: Morality. *Proceedings of the National Academy of Sciences*, *107*(Suppl. 2), 9015–9022.

Ayer, A. J. (1936/1946). *Language, Truth and Logic*. New York: Dover Publications.

Azzouni, J. (2008). A Cause for Concern: Standard Abstracta and Causation. *Philosophia Mathematica 16* (3), 397-401.

Beatty, J. (1993). The Evolutionary Contingency Thesis. In G. Wolters & J. G. Lennox (Eds.), *Concepts, Theories, and Rationality in the Biological Sciences: Pittsburgh-Konstanz Series in the Philosophy and History of Science* (pp. 45-82). Pittsburgh, P.A.: University of Pittsburgh Press.

Becker, K. (2008). Epistemic Luck and the Generality Problem. *Philosophical Studies, 139*(3), 353-366.

Becker, K. (2012). Methods and how to Individuate them. In K. Becker & T. Black (Eds.), *The Sensitivity Principle in Epistemology* (pp. 81-100). Cambridge: Cambridge University Press.

Bedke, M. (2008). Ethical Intuitions: What They are, What They are not, and How They Justify. *American Philosophical Quarterly, 45*(3), 253-269.

Bedke, M. (2009). Intuitive Non-Naturalism Meets Cosmic Coincidence. *Pacific Philosophical Quarterly, 90*(2), 188-209.

Bedke, M. (2014). No Coincidence?. In R. Shafer-Landau (Ed.), *Oxford Studies in Metaethics* (Vol. 9, pp. 102-125). Oxford: Oxford University Press.

Bell, D., & Hart, W. D. (1979). The Epistemology of Abstract Objects: Access and Inference. *Proceedings of the Aristotelian Society, Supplementary Volumes, 53*, 135-165.

Bennett, J. (2016). Accountability (II). In P. Russell & M. McKenna (Eds.), *Free Will and Reactive Attitudes: Perspectives on P.F. Strawson's Freedom and Resentment* (pp. 47-68). Oxon; New York: Routledge.

Beran, M. J. (2008). The Evolutionary and Developmental Foundations of Mathematics. *PLoS Biology, 6*(2), 221-223.

Beran, M. J. (2009). Chimpanzees as Natural Accountants. *Human evolution, 24*(3),

183-196.
Bernstein, S. (2019). Moral Luck and Deviant Causation. *Midwest Studies in Philosophy 43*(1), 151-161.
Birdsell, J. B. (1968). Some Predictions for the Pleistocene Based on Equilibrium Systems Among Recent Hunter-Gatherers. In R. B. Lee & I. DeVore (Eds.), *Man the Hunter* (pp. 229–240). New York: Aldine de Gruyter.
Björklund, F., Björnsson, G., Eriksson, J., Olinder, R. F., Strandberg, C. (2012). Recent Work on Motivational Internalism. *Analysis, 72*(1), 124-137.
Björnsson, G., & Finlay, S. (2010). Metaethical Contextualism Defended. *Ethics, 121*(1), 7-36.
Blackburn, S. (1971). Moral Realism. In J. Casey (Ed.), *Morality and Moral Reasoning* (pp. 101-124). London: Methuen.
Blackburn, S. (1984). *Spreading the Word: Groundings in the Philosophy of Language*. Oxford: Clarendon.
Blackburn, S. (1988). Attitudes and Contents. *Ethics, 98*(3), 501–517.
Blackburn, S. (1993). *Essays in Quasi-Realism*. New York: Oxford University Press.
Blackburn, S. (1996). Securing the Nots: Moral Epistemology for the Quasi-Realist. In W. Sinnott-Armstrong & M. Timmons (Eds.), *Moral Knowledge?: New Readings in Moral Epistemology* (pp. 82-100). New York: Oxford University Press.
Blackburn, S. (1998). *Ruling Passions: A Theory of Practical Reasoning*. New York: Oxford University Press.
Blackburn, S. (2005). Quasi-Realism no Fictionalism. In M. E. Kalderon (Ed.), *Fictionalism in Metaphysics* (pp. 322-338). New York: Oxford University Press.
Blackburn, S. (2006). Antirealist Expressivism and Quasi-Realism. In D. Copp (Ed.), *The Oxford Handbook of Ethical Theory* (pp. 146-162). New York: Oxford University Press.
Blackburn, S. (2009). Truth and A Priori Possibility: Egan's Charge Against Quasi-Realism. *Australasian Journal of Philosophy*, 87(2), 201-213.
Blackmore, S. (1999). *The Meme Machine*. New York: Oxford University Press.
Bogardus, T. (2016). Only All Naturalists Should Worry About Only One Evolutionary Debunking Argument. *Ethics, 126*(3), 636-661.
Bogardus, T. (2014). Knowledge Under Threat. *Philosophy and Phenomenological Research, 88*(2), 289-313.
Boghossian, P. (1994). Analyticity and Conceptual Truth. *Philosophical Issues, 5*, 117-131.
Boghossian, P. (1997). Analyticity. In B. Hale & C. Wright (Eds.), *A Companion to the Philosophy of Language,* (pp. 331-368). Oxford: Blackwell.
Boyd, R. (2007). How to be a Moral Realist. In Russ Shafer-Landau & Terence Cuneo (Eds.), *Foundations of Ethics: An Anthology* (pp. 163-185). Malden, M.A.: Blackwell Publishing.
Boyd, R., & Richerson, P. J. (1985). *Culture and the Evolutionary Process*. Chicago: University of Chicago Press.
Boyd, R., & Richerson, P. J. (1996). Why Culture is Common, but Cultural Evolution is Rare. *Proceedings of The British Academy, 88*, 77–93.
Boyd, R., & Richerson, P. J. (2005). *The Origin and Evolution of Cultures*. New York: Oxford University Press.
Bowles, S., & Gintis, H. (2011). *A Cooperative Species: Human Reciprocity and Its Evolution*. Princeton, N.J.; Oxford: Princeton University Press.

Boysen, S. T. (1993). Counting in Chimpanzees: Nonhuman Principles and Emergent Properties of Number. In S. T. Boysen & E. J. Capaldi (Eds.), C*omparative Cognition and Neuroscience. The Development of Numerical Competence: Animal and Human Models* (pp. 39–59). Hillsdale, N.J.,: Lawrence Erlbaum Associates.
Brandt, R. B. (1979). *A Theory of the Good and the Right*. Oxford: Clarendon Press.
Brandon, R. N. (1978). Adaptation and Evolutionary Theory. *Studies in History and Philosophy of Science Part A, 9*(3), 181-206.
Braddon-Mitchell, D., & Nola R. (2009). Introducing the Canberra Plan. In D. Braddon-Mitchell & R. Nola (Eds.), *Conceptual Analysis and Philosophical Naturalism* (pp. 1-20). Cambridge, M.A.: MIT Press.
Brandon, R. N. (1990). *Adaptation and Environment*. Princeton, N.J.: Princeton University Press.
Brink, D. O. (1989). *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.
Callard, B. (2007). The Conceivability of Platonism. *Philosophia Mathematica 15*(3), 347–56.
Carlin, J. L. (2011). Mutations are the Raw Materials of Evolution. *Nature Education Knowledge, 3*(10), 10.
Casullo, A. (2018). Pollock and Sturgeon on Defeaters. *Synthese 195*(7), 2897-2906.
Charlow, N. (2014). The Problem with the Frege-Geach Problem. *Philosophical Studies, 167*(3), 635-665.
Christensen, D. (2010). Higher-Order Evidence. *Philosophy and Phenomenological Research, 81*(1), 185-215.
Clarke, E. (2018). Adaptation, Multilevel Selection, and Organismality: A Clash of Perspectives. In Richard Joyce (Ed.), *The Routledge Handbook of Evolution and Philosophy* (pp. 35-48). Oxon; New York: Routledge.
Clarke-Doane, J. (2012). Morality and Mathematics: The Evolutionary Challenge. *Ethics, 122*(2), 313-340.
Clarke-Doane, J. (2014). Moral Epistemology: The Mathematics Analogy. *Noûs, 48*(2), 238-255.
Clarke-Doane, J. (2015). Justification and Explanation in Mathematics and Morality *Oxford Studies in Metaethics* (Vol. 10, pp. 80-103). Oxford: Oxford University Press.
Clarke-Doane, J. (2016). Debunking and Dispensability. In U. D. Leibowitz & N. Sinclair (Eds.), *Explanation in Ethics and Mathematics: Debunking and Dispensability* (pp. 23-35). Oxford: Oxford University Press.
Clarke-Doane, J. (2017a). Debunking Arguments: Mathematics, Logic, and Modal Security. In M. Ruse & R. J. Roberts (Eds.), *The Cambridge Handbook of Evolutionary Ethics* (pp. 202-209). Cambridge: Cambridge University Press.
Clarke-Doane, J. (2017b). What Is the Benacerraf Problem?. In F. Pataut (Ed.), *New Perspectives on the Philosophy of Paul Benacerraf: Truth, Objects, Infinity* (pp.17-43). Dordrecht: Springer.
Cohen, S. (1984). Justification and Truth. *Philosophical Studies, 46*(3), 279-295.
Collin, J. H. (2018). Towards an Account of Epistemic Luck for Necessary Truths. *Acta Analytica, 33*(4), 483-504.
Collins, J., Hall, N., & Paul, L. A. (2014). Counterfactuals and Causation: History, Problems, and Prospects. In J. Collins, N. Hall, and L. A. Paul (Eds.), *Causation and Counterfactuals* (pp. 1-58). Cambridge, M.A.: MIT Press.
Comesaña, J. (2005). Unsafe Knowledge. *Synthese, 146* (3), 395-404.

Confucius. ([2003]). *Confucius Analects* (E. Slingerland Trans.). Indianapolis, I.N.: Hackett Publishing.
Copp, D. (2001). *Morality, Normativity, and Society*. New York: Oxford University Press.
Copp, D. (2008). Darwinian Skepticism about Moral Realism. *Philosophical Issues, 18*(1), 186–206.
Crow, D. (2016). Causal Impotence and Evolutionary Influence: Epistemological Challenges for Non-Naturalism. *Ethical Theory and Moral Practice, 19*(2), 379-395.
Cummins, R. (1975). Functional Analysis. *Journal of Philosophy, 72*(20), 741-765.
Cuneo, T. (2007). Recent Faces of Moral Nonnaturalism. *Philosophy Compass, 2*(6), 850-879
Cuneo, T., & Shafer-Landau, R. (2014). The Moral Fixed Points: New Directions for Moral Nonnaturalism. *Philosophical Studies, 171*(3), 399-443.
Daly, C. (2008). Fictionalism and the Attitude. *Philosophical Studies, 139*(3), 423-440.
Daly, C. (2010). *An Introduction to Philosophical Methods*. Peterborough: Broadview.
Dancy, J. (1985). *Introduction to Contemporary Epistemology*. Oxford: Blackwell.
Darwin, C. (1876/2009). *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. New York: Cambridge University Press.
Darwin, C. (1871/2009). *The Descent of Man and Selection in Relation to Sex: Vol. 1*. New York: Cambridge University Press.
Das, R. (2016). Evolutionary Debunking of Morality: Epistemological or Metaphysical?. *Philosophical Studies*, *173*(2), 417-435.
Dawkins, R. (1976). *The Selfish Gene*. Oxford: Oxford University Press.
Dawkins, R. (1982). *The Extended Phenotype: The Gene as the Unit of Selection*. New York: Oxford University Press.
Dawkins, R. (1986/2006). *The Blind Watchmaker*. London: Penguin.
Dawkins, R. (2009). *The Greatest Show on Earth: The Evidence for Evolution*. New York: Free Press.
Dennett, D. (1995). *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. London: Penguin.
Dennett, D. (2001). The Evolution of Culture. *The Monist*, *84*(3), 305–324.
DeRose, K. (1995). Solving the Skeptical Problem. *The Philosophical Review, 104*(1), 1-52.
Dockstader, J. (2020a). Reactionary Fictionalism. *Southern Journal of Philosophy, 58*(2), 238-263.
Dockstader, J. (2020b). Reactionary Moral Fictionalism. *Philosophia, 48*(2), 519-534.
Eastwick, P. W. (2009). Beyond the Pleistocene: Using Phylogeny and Constraint to Inform the Evolutionary Psychology of Human Mating. *Psychological Bulletin*, *135*(5), 794–821.
Easwaran, K. (2015). Rebutting and Undercutting in Mathematics. *Philosophical Perspectives, 29*(1), 146-162.
Egan, A. (2007). Quasi-Realism and Fundamental Moral Error. *Australasian Journal of Philosophy*, 85(2), 205-219.
Enoch, D. (2007). An Outline of an Argument for Robust Metanormative Realism. In R. Shafer-Landau (Ed.), *Oxford Studies in Metaethics* (Vol. 2, pp. 21-50). Oxford: Oxford University Press.

Enoch, D. (2010). The Epistemological Challenge to Metanormative Realism: How Best to Understand it, and How to Cope with it. *Philosophical Studies, 148*(3), 413-438.
Enoch, D. (2011). *Taking Morality Seriously: A Defense of Robust Realism*. New York: Oxford University Press.
Enoch, D., & McPherson, T. (2017). What Do you Mean "This isn't the Question"?. *Canadian Journal of Philosophy 47*(6), 820-840.
Evers, D., & Streumer, B. (2016). Are the Moral Fixed Points Conceptual Truths?. *Journal of Ethics and Social Philosophy, 10*(1), 1-9.
Feldman, R. (2003). *Epistemology*. Upper Saddle River, N.J.: Prentice Hall.
Fine, K. (2002). Varieties of Necessity. In T. S. Gendler & J. Hawthorne (Eds.), *Conceivability and Possibility* (pp. 253-281). Oxford: Oxford University Press.
Finlay, S. (2007). Four Faces of Moral Realism. *Philosophy Compass, 2*(6), 820-849.
Finlay, S. (2014). *Confusion of Tongues: A Theory of Normative Language*. New York: Oxford University Press.
Fisher, A. (2011). *Metaethics: An Introduction*. Durham: Acumen Press.
FitzPatrick, W. (2014). Why There Is No Darwinian Dilemma for Ethical Realism. In M. Bergmann & P. Kain (Eds.), *Challenges to Moral and Religious Belief: Disagreement and Evolution* (pp. 237-254). Oxford: Oxford University Press.
FitzPatrick, W. (2015). Debunking Evolutionary Debunking of Ethical Realism. *Philosophical Studies, 172*(4), 883-904.
Fodor, J. (1998). *Concepts: Where Cognitive Science Went Wrong*. New York: Oxford University Press.
Foot, P. (1958). Moral Beliefs. *Proceedings of the Aristotelian Society, 59*, 83-104.
Foot, P. (1967). The Problem of Abortion and the Doctrine of Double Effect. *Oxford Review,* 5, 5-15.
Fraser, B. J. (2010). Adaptation, Exaptation, By-Products, and Spandrels in Evolutionary Explanations of Morality. *Biological Theory, 5*(3), 223–227.
Fraser, B. J. (2014). Evolutionary Debunking Arguments and The Reliability of Moral Cognition. *Philosophical Studies, 168*(2), 457-473.
Garson, J. (2019). There Are No Ahistorical Theories of Function. *Philosophy of Science, 86*(5), 1146-1156.
Gaus, G. (2010). The Demands of Impartiality and the Evolution of Morality. In B. Feltham & J. Cottingham (Eds.), *Partiality and Impartiality: Morality, Special Relationships, and the Wider World* (pp. 42-64). New York: Oxford University Press.
Gaus, G. (2011). *The Order of Public Reason*. New York: Cambridge University Press.
Geach, P. T. (1960). Ascriptivism. *Philosophical Review, 69*(2), 221-225.
Geach, P. T. (1965). Assertion. *Philosophical Review, 74*(4), 449-465.
Gettier, E. (1963). Is Justified True Belief Knowledge?. *Analysis, 23*(6), 121-123.
Gibbard, A. (1985). Moral Judgment and the Acceptance of Norms. *Ethics, 96*(1), 5-21.
Gibbard, A. (1990). *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Oxford: Clarendon Press.
Gibbard, A. (2003). *Thinking How to Live*. Cambridge, Cambridge, M.A.; London: Harvard University Press.
Godfrey-Smith, P. (1994). A Modern History Theory of Functions. *Noûs, 28*(3), 344-362.
Goldman, A. (1976). Discrimination and Perceptual Knowledge. *Journal of Philosophy 73*(20), 771-791.

Gould, S. J., & Lewontin, R. C. (1979). The Spandrels of San Marco and the Panglossian Paradigm: a Critique of the Adaptationist Programme. *Proceedings of the Royal Society B: Biological Sciences*, *205*(1), 581–598.
Gould, S. J., & Vrba, E. S. (1982). Exaptation – a Missing Term in the Science of Form. *Paleobiology*, *8*(1), 4–15.
Gould, S. J. (1991). *Bully for Brontosaurus: Reflections in Natural History*. New York; London: W. W. Norton & Company.
Greene, J. D. (2002). *The Terrible, Horrible, No Good, Very Bad Truth About Morality, and What to Do About It* (Doctoral dissertation). Princeton University, Princeton, N.J..
Greco, J. (2010). *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity*. Cambridge: Cambridge University Press.
Greco, J. (2012). Better Safe than Sensitive. In T. Black & K. Becker (Eds.), *The Sensitivity Principle in Epistemology* (pp. 193-206). Cambridge: Cambridge University Press.
Greene, J. D. (2016). Solving the Trolley Problem. In J. Sytsma & W. Buckwalter (Eds.), *A Companion to Experimental Philosophy* (pp. 175–178). Hoboken, N.J.: Wiley.
Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science, 293*(5537), 2105–2108.
Griffin, J. (2008). *On Human Rights*. New York: Oxford University Press.
Gruber, T., Luncz, L., Mörchen, J., Schuppli, C., Kendal, R. L., & Hockings, K. (2019). Cultural Change in Animals: a Flexible Behavioural Adaptation to Human Disturbance. *Palgrave Communications, 5*(64), 1-9.
Grundmann, T. (2011). Defeasibility Theory. In S. Bernecker & D. Pritchard (Eds.), *The Routledge Companion to Epistemology* (pp. 156-166). Oxon; New York: Routledge.
Haidt, J. (2001). The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgment. *Psychological Review*, *108*(4), 814–834.
Haidt, J., & Hersh, M. A. (2001). Sexual Morality: The Cultures and Emotions of Conservatives and Liberals. *Journal of Applied Psychology*, *31*(1), 191–221.
Hamilton, W.D. (1963). The Evolution of Altruistic Behaviour. *The American Naturalist, 97*(896), 354-356.
Hare, R. M. (1952/1991). *The Language of Morals*. Oxford: Oxford University Press.
Harman, G. (1988). Ethics and Observation. In G. Sayre-McCord (Ed.), *Essays on Moral Realism* (pp. 119-126). Ithaca, N.Y.: Cornell University Press.
Harman, G. (2015). Moral Relativism is Moral Realism. *Philosophical Studies, 172*(4), 855-863.
Hayward, M. K. (2018). Non-Naturalist Moral Realism and the Limits of Rational Reflection. *Australasian Journal of Philosophy 96*(4):724-737.
Heathwood, C. (2012). Could Morality have a Source?. *Journal of Ethics and Social Philosophy 6*(2), 1-19.
Henrich, J., & Boyd, R. (2002). On Modeling Cognition and Culture: Why Cultural Evolution Does Not Require Replication of Representations. *Journal of Cognition and Culture*, *2*(2), 87–112.
Henrich, J., Boyd, R., & Richerson, P. J. (2008). Five Misunderstandings About Cultural Evolution. *Human Nature*, *19*(2), 119–137.
Hirvelä, J. (2019). Global Safety: How to Deal with Necessary Truths. *Synthese, 196*

(3), 1167-1186.
Hooker, B. (2002). *Ideal Code, Real World: A Rule-Consequentialist Theory of Morality*. New York: Oxford University Press.
Huemer, M. (2005). *Ethical Intuitionism*. New York: Palgrave Macmillan Press.
Huemer, M. (2015). The Failure of Analysis and the Nature of Concepts. In C. Daly (Ed.), *The Palgrave Handbook of Philosophical Methods* (pp. 51-76). Basingstoke: Palgrave MacMillan.
Huemer, M. (2016). A Liberal Realist Answer to Debunking Skeptics: The Empirical Case for Realism. *Philosophical Studies, 173*(7), 1983–2010.
Hume, D. (1748/2007). *An Enquiry Concerning Human Understanding*, P. Millican (Ed.). New York: Oxford University Press.
Hung, C-H., & Tse, C. Y. P. (forthcoming). The Frege-Geach Problem and Blackburn's Expressivism. *Philosophia*.
Ingram, S. (2017). I Can't Relax! You're Driving me Quasi!. *Pacific Philosophical Quarterly, 98*(3), 490-510.
Jablonka, E., & Lamb, M. J. (2005). *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioural, and Symbolic Variation in the History of Life*. Cambridge, M.A.; London: MIT Press.
Jackson, F. (1998). *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Oxford University Press.
Jackson, F. (2008). The Argument from the Persistence of Moral Disagreement. In R. Shafer-Landau (Ed.), *Oxford Studies in Metaethics* (Vol. 3, pp. 75-86). Oxford: Oxford University Press.
Jackson, F., & Pettit, P. (1998). A Problem for Expressivism. *Analysis, 58*(4), 239-251.
James, S. M. (2011). *An Introduction to Evolutionary Ethics*. Chichester: Blackwell.
Janvid, M. (2017). Defeater Goes External. *Philosophia 45*(2), 701-715.
Johannsen, W. (1911). The Genotype Conception of Heredity. *The American Naturalist, 45*(531), 129-159.
Jonas, S. (2017). Access Problems and Explanatory Overkill. *Philosophical Studies 174* (11), 2731-2742.
Jong, J., & Visala, A. (2014). Evolutionary Debunking Arguments against Theism, Reconsidered. *International Journal for Philosophy of Religion 76*(3), 243-258.
Joyce, R. (2000). Darwinian Ethics and Error. *Biology and Philosophy*, *15*(5), 713–732.
Joyce, R. (2001). *The Myth of Morality*. Cambridge: Cambridge University Press.
Joyce, R. (2005). Moral Fictionalism. In M. E. Kalderon (Ed.), *Fictionalism in Metaphysics* (pp. 287-313). New York: Oxford University Press.
Joyce, R. (2006). *The Evolution of Morality*. Cambridge, M.A.: MIT Press.
Joyce, R. (2016a). Evolution, Truth-tracking, and Moral Skepticism. In *Essays in Moral Skepticism* (pp. 142–158). Oxford: Oxford University Press.
Joyce, R. (2016b). Reply: Confessions of a modest debunker. In U. D. Leibowitz & N. Sinclair (Eds.), *Explanation in Ethics and Mathematics: Debunking and Dispensability* (pp. 124-144). Oxford: Oxford University Press.
Joyce, R. (2017). Human Morality: From an Empirical Puzzle to a Metaethical Puzzle. In M. Ruse & R. J. Richards (Eds.), *The Cambridge Handbook of Evolutionary Ethics* (pp. 101–113). Cambridge: Cambridge University Press.
Kagan, S. (1989). *The Limits of Morality*. New York: Oxford University Press.
Kahane, G. (2011). Evolutionary Debunking Arguments. *Noûs, 45*(1), 103-125.
Kauppinen, A. (2013). A Humean Theory of Moral Intuition. *Canadian Journal of*

Philosophy, 43(3), 360-381.
Kauppinen, A. (2015). Intuition and Belief in Moral Motivation. In G. Björnsson, C. Strandberg, R. F. Olinder, J. Eriksson, & F. Björklund (Eds.), *Motivational Internalism* (pp. 237–259). New York: Oxford University Press.
Kelly, T. (2010). Peer Disagreement and Higher-Order Evidence. In R. Feldman & T. A. Warfield (Eds.), *Disagreement* (pp. 111-174). Oxford: Oxford University Press.
Kelp, C. (2009). Knowledge and Safety. *Journal of Philosophical Research, 34*, 21-31.
Killin, A. (2018). Music and Human Evolution: Philosophical Aspects. In R. Joyce (Ed.), *The Routledge Handbook of Evolution and Philosophy* (pp. 372-386). Oxon; New York: Routledge.
Killoren, D. (2016). Why Care About Moral Fixed Points?. *Analytic Philosophy, 57*(2), 165-173.
Kitcher, P. (2006). Ethics and Evolution: How to get Here from There. In F. de Waal (Ed.), *Primates and Philosophers: How Morality Evolved* (pp. 120-139). Princeton, N.J.: Princeton University Press.
Klein, R-T. (2014). Where there are Internal Defeaters, there are "Confirmers". *Synthese, 191*(12), 2715-2728.
Klenk, M. (2017). Old Wine in New Bottles: Evolutionary Debunking Arguments and the Benacerraf-Field Challenge. *Ethical Theory and Moral Practice, 20*(4), 781-795.
Korman, D. Z., & Locke, D. (forthcoming). Against Minimalist Responses to Moral Debunking Arguments. *Oxford Studies in Metaethics*.
Köhler, S. (2012). Expressivism, Subjectivism and Moral Disagreement. *Thought: A Journal of Philosophy, 1*(1), 71-78.
Kripke, S. (2011). Nozick on Knowledge. In *Philosophical Troubles: Collected Papers* (Vol. 1, pp. 162-224). New York: Oxford University Press.
Kuper, A. (2000). If Memes are the Answer, What is the Question?. In R. Aunger (Ed.), *Darwinizing Culture: The Status of Memetics as a Science* (pp. 175–188). Oxford: Oxford University Press.
Laland, K. N., & Janik, V. M. (2006). The Animal Cultures Debate. *Trends in Ecology and Evolution, 21*(10), 542-547.
Landini, G. (2011). *Russell*. Oxon; New York: Routledge.
Laskowski, N. & Finlay, S. (2017). Conceptual Analysis in Metaethics. In T. McPherson & D. Plunkett (Eds.), *The Routledge Handbook of Metaethics* (pp. 536-551). London: Routledge.
Lasonen-Aarnio, M. (2014). Higher-Order Evidence and the Limits of Defeat. *Philosophy and Phenomenological Research, 88*(2), 314-345.
Leibowitz, U. D., & Sinclair, N. (2017). Evolution and the Missing Link (in Debunking Arguments). In M. Ruse & R. J. Roberts (Eds.), *The Cambridge Handbook of Evolutionary Ethics* (pp. 210-225). Cambridge: Cambridge University Press.
Lennertz, B. (forthcoming). Noncognitivism and the Frege-Geach Problem in Formal Epistemology. *Philosophy and Phenomenological Research*.
Levy, A., & Levy, Y. (2016). The Debunking Challenge to Realism: How Evolution (Ultimately) Matters. *Journal of Ethics and Social Philosophy*, *11*(1), 1–8.
Lewens, T. (2015). *Cultural Evolution: Conceptual Challenges*. Oxford: Oxford University Press.
Lewis, D. (1973). Causation. *The Journal of Philosophy, 70*(17), 556-567.
Lewis, D. (1983). *Philosophical Papers* (Vol. 1). New York: Oxford University Press.

Lewis, D. (2005). Quasi-Realism is Fictionalism. In M. E. Kalderon (Ed.), *Fictionalism in Metaphysics* (pp. 314-321). New York: Oxford University Press.
Lieberman, D. (2008). Moral Sentiments Relating to Incest: Discerning Adaptations from By-products. In W. Sinnott-Armstrong (Ed.), *Moral Psychology: Volume 1: The Evolution of Morality: Adaptations and Innateness* (pp. 165–190). Cambridge, M.A.; London: MIT Press.
Lillehammer, H. (2011). The Epistemology of Ethical Intuitions. *Philosophy, 86*(336), 175-200.
Lorre, C. (Writer), Prady, B. (Writer), & Cendrowski, M. (Director) (2007). The Roommate Transmogrification [Television series episode]. In Lorre, C. (Executive producer), & Prady, B. (Executive producer), *The Big Bang theory*. New York: Columbia Broadcasting System.
Lutz, M. (2015). *The Case for Moral Skepticism.* (Doctoral Dissertation), The University of Southern California, Los Angeles, C.A..
Lutz, M. (2018). What Makes Evolution a Defeater?. *Erkenntnis, 83*(6), 1105-1126.
Mackie, J. L. (1977). *Ethics: Inventing Right and Wrong*. London: Penguin Books.
Matthews, G. B. (2005). The Ontological Argument. In W. E. Mann (Ed.), *The Blackwell Guide to the Philosophy of Religion* (pp. 81-102). Malden, M.A.: Blackwell.
Mayr, E. (1961). Cause and Effect in Biology. *Science*, *134*(3489), 1501–1506.
McGrath, M. (forthcoming). Undercutting Defeat: When it Happens and Some Implications for Epistemology. In J. Brown and M. Simionescu (Eds.), *Reasons, Justifications and Defeat*. Oxford: Oxford University Press.
McHugh, C. (2010). Self-Knowledge and the KK Principle. *Synthese, 173*(3), 231-257.
McKay, R., & Dennett, D. (2009). Our Evolving Beliefs about Evolved Misbelief. *Behavioural and Brain Sciences, 32*(6), 541-561.
McMahan, J. (2013). Moral Intuition. In H. LaFollette & I. Persson (Eds.), *The Blackwell Guide to Ethical Theory* (pp. 103-120). Chicester: Wiley-Blackwell.
McPherson, T. (2012). Ethical Non-Naturalism and the Metaphysics of Supervenience. In R. Shafer-Landau (Ed.), *Oxford Studies in Metaethics* (Vol. 7, pp. 205-234). Oxford: Oxford University Press.
McPherson, T. (2020). *Epistemology and Methodology in Ethics*. Cambridge: Cambridge University Press.
Menzies, P. (1989). Probabilistic Causation and Causal Processes: A Critique of Lewis. *Philosophy of Science, 56*(4), 642–663.
Mesoudi, A., & Danielson, P. (2008). Ethics, Evolution and Culture. *Theory in Biosciences*, *127*(3), 229–240.
Mesoudi, A., & Thornton, A. (2018). What is Cumulative Cultural Evolution?. *Proceedings of the Royal Society B, 285*(1880), 1-8.
Mill, J. S. (1861/2003). Utilitarianism. In M. Warnock (Ed.), *Utilitarianism and On Liberty: Including Mill's 'Essay on Bentham' and Selections from the Writings of Jeremy Bentham and John Austin* (2nd ed., pp. 181–235). Oxford: Blackwell Publishing.
Miller, A. (2003). *An Introduction of Contemporary Metaethics*. Cambridge: Polity Press.
Mogensen, A. (2015). Evolutionary debunking arguments and the proximate/ultimate distinction. *Analysis, 75*(2), 196-203.
Mogensen, A. (2016). Do Evolutionary Debunking Arguments Rest on a Mistake About

Evolutionary Explanations?. *Philosophical Studies, 173*(7), 1799-1817.
Moon, A. (2017). Debunking Morality: Lessons from the EAAN Literature. *Pacific Philosophical Quarterly, 98*(S1), 208-226.
Moore, G. E. (1899). The Nature of Judgment. *Mind, 8*(30), 176-193.
Morin, O. (2016a). Reasons to be Fussy about Cultural Evolution. *Biology and Philosophy*, *31*(3), 447–458.
Morin, O. (2016b). *How Traditions Live and Die*. New York: Oxford University Press.
Mulgan, T. (2006). *Future People: A Moderate Consequentialist Account of Our Obligations to Future Generations*. New York: Oxford University Press.
Nagel, J. (2013). Defending the Evidential Value of Epistemic Intuitions: A Reply to Stich. *Philosophy and Phenomenological Research, 86*(1), 179-199.
Nichols, S. (2014). Process Debunking and Ethics. *Ethics, 124*(4), 727-749.
Nolan, D., Restall, G., & West, C. (2005). Moral Fictionalism Versus the Rest. *Australasian Journal of Philosophy, 83*(3), 307-330.
Nowak, M. A. (2006). Five Rules for the Evolution of Cooperation. *Science, 314*(5805), 1560-1563.
Nozick, R. (1981). *Philosophical Explanations*. Cambridge, M.A.: Harvard University Press.
Okasha, S. (2013). On a Flawed Argument against the KK Principle. *Analysis, 73*(1), 80-86.
Orr, H. A. (2005). The Genetic Theory of Adaptation: a Brief History. *Nature Reviews Genetics, 6*, 119-127.
Parfit, D. (1984). *Reasons and Persons*, Oxford: Oxford University Press.
Parfit, D. (2011). *On What Matters Vol. 2*, S. Scheffler (Ed.). Oxford: Oxford University Press.
Park, S. (2019). Can Mathematical Objects be Causally Efficacious?. *Inquiry, 62*(3), 247-255.
Plantinga, A. (2000). *Warranted Christian Belief*. New York: Oxford University Press.
Plato. ([1997]). Theaetetus (M. J. Levett, Trans., M. Burnyeat, Rev.). In J. M. Cooper & D. S. Hutchinson (Eds.), *Complete Works* (pp. 157-234). Indianapolis, I.N.: Hackett Publishing
Plato. ([1997]). Meno (G.M.A. Grube, Trans.). In J. M. Cooper & D. S. Hutchinson (Eds.), *Complete Works* (pp. 870-897). Indianapolis, I.N.: Hackett Publishing
Pollock, J. L. (1974). *Knowledge and Justification*. Princeton, N.J.: Princeton University Press.
Pollock, J. L. (1987). Defeasible Reasoning. *Cognitive Science, 11*(4), 481-518.
Prinz, J. (2006). The Emotional Basis of Moral Judgments. *Philosophical Explorations 9* (1), 29-43.
Prinz, J. (2007). *The Emotional Construction of Morals*. New York: Oxford
University Press.
Prinz, J. (2015). An Empirical Case for Motivational Internalism. In G. Björnsson, C. Strandberg, R. F. Olinder, J. Eriksson, & F. Björklund (Eds.), *Motivational Internalism* (pp. 61–84). New York: Oxford University Press.
Pritchard, D. (2005). *Epistemic Luck*. New York: Oxford University Press.
Pritchard, D. (2008). Sensitivity, Safety, and Anti-Luck Epistemology. In J. Greco (Ed.), *The Oxford Handbook of Skepticism* (pp. 437-455). Oxford: Oxford University Press.
Pritchard, D. (2012a). In Defence of Modest Anti-Luck Epistemology. In T. Black & K. Becker (Eds.), *The Sensitivity Principle in Epistemology* (pp. 173-192). Cambridge:

Cambridge University Press.
Pritchard, D. (2012b). Anti-Luck Virtue Epistemology. *Journal of Philosophy, 109* (3), 247-279.
Pritchard, D. (2014). Anti-luck Epistemology and the Gettier Problem. *Philosophical Studies, 172*(1), 93-111.
Pritchard, D. (2016). *Epistemology* (2nd ed.). Basingstoke: Palgrave Macmillan.
Pritchard, D. (2018). Anti-Luck Virtue Epistemology and Epistemic Defeat. *Synthese 195*(7), 3065-3077.
Putnam, H. (1983). Two dogmas' revisited. In *Realism and Reason: Philosophical Papers Vol. 3* (pp. 87-97). Cambridge: Cambridge University Press.
Pust, J. (2019). Intuition. In E. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Available from: https://plato.stanford.edu/archives/sum2019/entries/intuition/ (last accessed 12 July, 2020).
Railton, P. (1986). Moral realism. *The Philosophical Review, 95*(2), 163–207.
Ramachandran, M. (1997). A Counterfactual Analysis of Causation. *Mind, 106*(422), 263-277.
Ramsey, W. (2019). Intuitions as Evidence Facilitators. *Metaphilosophy, 50*(1-2), 76-99.
Rawls, J. (1951). Outline of a Decision Procedure for Ethics. *The Philosophical Review. 60*(2), 177-197.
Rawls, J. (1971/1999). *A Theory of Justice (rev. ed.)*. Cambridge, M.A.: Belknap Press.
Richerson, P. (2011). Evolution: Not So Selfish. *Nature, 476*, 29-30.
Ridge, M. (2019). Moral Non-Naturalism. In E. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. Available from: https://plato.stanford.edu/archives/fall2019/entries/moral-non-naturalism/ (last accessed 12 July, 2020).
Roland, J., & Cogburn, J. (2011). Anti-Luck Epistemologies and Necessary Truths. *Philosophia, 39*(3), 547-561
Rolfe, R. D., Hentges, D. J., Campbell, B. J., & Barrett, J. T. (1978). Factors Related to the Oxygen Tolerance of Anaerobic Bacteria. *Applied and Environmental Microbiology*, *36*(2), 306–313.
Rosen, G. (forthcoming). What is Normative Necessity?. In M. Dumitru (Ed.), *Metaphysics, Meaning, and Modality: Themes from Kit Fine*. Oxford: Oxford University Press.
Rowe, W. L. (1978/2017). The Ontological Argument. In J. Feinberg & R. Shafer-Landau (Eds.) *Reason and Responsibility: Readings in Some Basic Problems of Philosophy* (pp. 36-46). Boston, M.A.: Cengage Learning.
Rowland, R. (2019). Local Evolutionary Debunking Arguments. *Philosophical Perspectives 33*(1), 170-199.
Ruse, M. (1995). *Evolutionary Naturalism*. Oxon: Routledge.
Russell, B. (1903/2010). *Principles of Mathematics*. Oxford: Routledge.
Sayre-McCord, G. (1988). Moral Theory and Explanatory Impotence. In G. Sayre-McCord (Ed.), *Essays on Moral Realism* (pp. 256-281). Ithaca, N.Y.: Cornell University Press.
Scanlon, T. (2014). *Being Realistic about Reasons*. Oxford: Oxford University Press.
Schaffer, J. (2001). Causes as Probability-Raisers of Processes. *Journal of Philosophy, 98*(2), 75–92.
Schantz, R. (2004). Introduction. In R. Schantz (Ed.), *The Externalist Challenge* (pp.

1-36). Berlin: Walter de Gruyter.
Schroeder, M. (2005). Realism and Reduction: The Quest for Robustness. *Philosophers' Imprint, 5*(1), 1-18.
Schroeder, M. (2007). *Slaves of Passions*. Oxford: Oxford University Press.
Schroeder, M. (2014). The Price of Supervenience. In *Explaining the Reasons we Share: Explanation and Expression in Ethics* (Vol. 1, pp.124-144). Oxford: Oxford University Press.
Schulz. A. (2013). Exaptation, Adaptation, and Evolutionary Psychology. *History and Philosophy of the Life Sciences, 35*(2), 193-212.
Scott-Phillips, T. C., Dickins, T. E., & West, S. A. (2011). Evolutionary Theory and the Ultimate–Proximate Distinction in the Human Behavioural Sciences. *Perspectives on Psychological Science*, *6*(1), 38–47.
Setiya, K. (2012). *Knowing Right from Wrong*. Oxford: Oxford University Press.
Shafer-Landau, R. (2003). *Moral Realism: A Defense*. Oxford: Oxford University Press.
Shafer-Landau, R. (2012). Evolutionary Debunking, Moral Realism and Moral Knowledge. *Journal of Ethics & Social Philosophy, 7*(1), 1-38.
Shafer-Landau, R. (2017). Moral Realism and Evolutionary Debunking Arguments. In M. Ruse & R. J. Roberts (Eds.), *The Cambridge Handbook of Evolutionary Ethics* (pp. 175-187). Cambridge: Cambridge University Press.
Sidgwick, H. (1907/1962). *The Method of Ethics*. London: Palgrave Macmillan.
Silk, A. (2013). Truth Conditions and the Meanings of Ethical Terms 1. In R. Shafer-Landau (Ed.), *Oxford Studies in Metaethics* (Vol. 8, pp. 195-222). Oxford: Oxford University Press.
Simmons, R., & Scheepers, L. (1996). Winning by a Neck: Sexual Selection in the Evolution of Giraffe. *The American Naturalist*, *148*(5), 771-786.
Sinclair, N. (2009). Recent Work in Expressivism. *Analysis, 69*(1), 136-147.
Sinclair, N. (2012). Metaethics, Teleosemantics and the Function of Moral Judgements. *Biology and Philosophy, 27*(5), 639-662.
Sinclair, N. (2018). Belief Pills and the Possibility of Moral Epistemology. In R. Shafer-Landau (Ed.), *Oxford Studies in Metaethics* (Vol. 13, pp. 98-122). Oxford: Oxford University Press.
Singer, P. (2005). Ethics and Intuitions. *The Journal of Ethics, 9*(3-4), 331-352.
Smart, J. J. C. (1973). An Outline of a System of Utilitarian Ethics. In J. J. C. Smart & B. Williams (Eds.), *Utilitarianism: For and Against* (pp. 1-76). Cambridge: Cambridge University Press.
Smearman, C. (2009). Second Wives Club: Mapping the Impact of Polygamy in US Immigration Law. *Berkeley Journal of International Law, 27*(2), 382-447.
Smith, M. (1994). *The Moral Problem*. Oxford: Blackwell Publishing.
Sober, E., & Wilson, D. S. (1998). *Unto others: The Evolution and Psychology of Unselfish Behaviour*. Cambridge, M.A.: Harvard University Press.
Sosa, E. (1999). How to defeat opposition to Moore?. *Philosophical Perspectives, 13*, 141-154.
Sperber, D. (2000). An Objection to the Memetic Approach to Culture. In R. Aunger (Ed.), *Darwinizing Culture: The Status of Memetics as a Science* (pp. 163-173). Oxford: Oxford University Press.
Sterelny K., & Griffiths P. (1999). *Sex and Death: An Introduction to Philosophy of Biology*. Chicago; London: The University of Chicago Press.

Sterelny, K. (2006). Memes Revisited. *The British Journal for the Philosophy of Science*, *57*(1), 145–165.
Stevenson, C. L. (1944). *Ethics and Language*. New Haven, C.T.: Yale University Press.
Stratton-Lake, P. (2016). Intuition, Self-Evidence, and Understanding. In R. Shafer-Landau (Ed.), *Oxford Studies in Metaethics* (Vol. 11, pp. 28-44). Oxford: Oxford University Press.
Strawson, P. F. (1962/2008). Freedom and Resentment. In *Freedom and Resentment And other essays* (pp. 1-28). Oxon; New York: Routledge.
Street, S. (2006). A Darwinian Dilemma for Realist Theories of Value. *Philosophical Studies, 127*(1), 109-166.
Street, S. (2008). Reply to Copp: Naturalism, Normativity, and the Varieties of Realism Worth Worrying About. *Philosophical Issues, 18*(1), 207-228.
Street, S. (2016). Objectivity and Truth: You'd Better Rethink It. In R. Shafer-Landau (Ed.), *Oxford Studies in Metaethics* (Vol. 11, pp. 293-334). Oxford: Oxford University Press.
Sturgeon, N. (1988). Moral Explanations. In G. Sayre-McCord (Ed.), *Essays on Moral Realism* (pp. 229-255). Ithaca, N.Y.: Cornell University Press.
Sturgeon, N. (2006). Ethical Naturalism. In D. Copp (Ed.), *The Oxford Handbook of Ethical Theory* (pp. 91-121). New York: Oxford University Press.
Sturgeon, S. (1993). The Gettier Problem. *Analysis, 53*(3), 156-164.
Sturgeon, S. (2014). Pollock on Defeasible Reasons. *Philosophical Studies, 169*(1), 105-118.
Sturgeon, S. (forthcoming). Undercutting Defeat and Edgington's Burglar. In L. Walters & J. Hawthorne (Eds.), *Conditionals, Paradox, and Probability: Themes from the Philosophy of Dorothy Edgington*. Oxford: Oxford University Press.
Suikkanen, J. (2016). Naturalism in Metaethics. In K. J. Clark (Ed.), *Blackwell Companion to Naturalism* (pp. 351-368). Malden, M.A.: Wiley-Blackwell.
Svavarsdottir, S. (1999). Moral Cognitivism and Motivation. *Philosophical Review*, *108*(2), 161–219.
Tersman, F. (2016). Explaining the Reliability of Moral Beliefs. In U. D. Leibowitz & N. Sinclair (Eds.), *Explanation in Ethics and Mathematics: Debunking and Dispensability* (pp. 38-57). Oxford: Oxford University Press.
Tiberius, V. (2015). *Moral Psychology: A Contemporary Introduction*. Oxon; New York: Routledge.
Thomson, J. J. (1985). The Trolley Problem. *The Yale Law Journal, 94*(6), 1395-1415.
Thomson, J. J. (1976). Killing, Letting Die, and the Trolley Problem. *The Monist, 59*(2), 204-217.
Thune, M. (2010). 'Partial Defeaters' and the Epistemology of Disagreement. *Philosophical Quarterly 60*(239), 355-372.
Toppinen, T. (2014). *Essays on Expressivism.* (Doctoral Dissertation), University of Helsinki, Helsinki.
Trivers. R. L. (1971). The Evolution of Reciprocal Altruism. *The Quarterly Review of Biology, 46*(1), 35-57.
Tye, M. (1991). *The Imagery Debate*. Cambridge, M.A.; London: MIT Press.
Vahid, H. (2009). *The Epistemology of Belief*. Basingstoke: Palgrave Macmillan.
Vahid, H. (2011). Externalism/Internalism. In S. Bernecker & D. Pritchard (Eds.), *The Routledge Companion to Epistemology* (pp. 144-155). Oxon; New York: Routledge.

Van Inwagen, P. (1997). Materialism and the Psychological-Continuity Account of Personal Identity. *Philosophical Perspectives 11*, 305-319.
Vavova, K. (2014). Debunking Evolutionary Debunking. In R. Shafer-Landau (Ed.), *Oxford Studies in Metaethics* (Vol. 9, pp. 76-101). Oxford: Oxford University Press.
Vavova, K. (2015). Evolutionary Debunking of Moral Realism. *Philosophy Compass, 10*(2), 104-116.
Väyrynen, P. (2013). *The Lewd, the Rude and the Nasty: A Study of Thick Concepts in Ethics*. New York: Oxford University Press.
Väyrynen, P. (2018). The Supervenience Challenge to Non-Naturalism. In T. McPherson & D. Plunkett (Eds.), *The Routledge Handbook of Metaethics* (pp. 170-184). New York: Routledge.
Wedgewood, R. (2007). *The Nature of Normativity*. New York: Oxford University Press.
Wheatley T., & Haidt J. (2005). Hypnotic Disgust Makes Moral Judgments More Severe. *Psychological Science, 16*(10), 780-4.
Wielenberg, E. (2010). On the Evolutionary Debunking of Morality. *Ethics, 120*(3), 441-464.
Wielenberg, E. (2014). *Robust Ethics: The Metaphysics and Epistemology of Godless Normative Realism*. Oxford: Oxford University Press.
Wielenberg, E. (2016). Ethics and Evolutionary Theory. *Analysis, 76*(4), 502-515.
Williams, M. (2001). *Problems of Knowledge: A Critical Introduction to Epistemology*. Oxford: Oxford University Press.
Williamson, T. (2007). *The Philosophy of Philosophy*. Oxford: Blackwell.
Williamson, T. (2011). Knowledge First Epistemology. In S. Bernecker & D. Pritchard (Eds.), *The Routledge Companion to Epistemology* (pp. 208-218). London: Routledge.
Wilson, A. (2018). Metaphysical Causation. *Noûs, 52*(4), 723-751.
Wilson, D. S. (1992). On the Relationship Between Evolutionary and Psychological Definitions of Altruism and Selfishness. *Biology and Philosophy*, *7*(1), 61–68.
Wong, D. (1984). *Moral Relativity*. Los Angles; London: University of California Press.
Woodruff, G., & Premack, D. (1981). Primative Mathematical Concepts in the Chimpanzee: Proportionality and Numerosity. *Nature, 293*, 568-570.
Woods, J. (2017). The Frege-Geach Problem. In T. McPherson & D. Plunkett (Eds.), *The Routledge Handbook of Metaethics* (pp. 226-242). New York: Routledge.
Woods, J. (2018). Mathematics, Morality, and Self-effacement. *Noûs, 52*(1), 47-68.
Wright, L. (1973). Functions. *Philosophical Review, 82*(2), 139-168.
Wright, L. (1976). *Teleological Explanations*. Berkeley: University of California Press.
Zhong, L. (2010). *A Non-Reductive Naturalist Approach to Moral Explanation.* (Doctoral Dissertation), The University of Michigan, Ann Arbor, M.I..