

KANTIAN CONSTITUTIVISM AS AN ALTERNATIVE TO MORAL REALISM

AN ABSTRACT

SUBMITTED ON THE THIRD DAY OF MAY 2022

TO THE DEPARTMENT OF PHILOSOPHY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

OF THE SCHOOL OF LIBERAL ARTS

OF TULANE UNIVERSITY

FOR THE DEGREE

OF

DOCTOR OF PHILOSOPHY

BY



Caner Turan

APPROVED: 

Oliver Sensen, Ph.D.
Director



Alison Denham, Ph.D.



Bruce Brower, Ph.D.

ABSTRACT

Moral realism is a metaethical theory according to which there are moral statements that are true independent of our attitudes toward them, and the facts of morality are discovered rather than created by humans. This view has the advantage of capturing our commitment to objective moral truths in our intuitive starting points. However, its commitment to stance-independence alienates us from moral truths in metaphysically and epistemically problematic ways. The aim of this dissertation is to explore alternatives to moral realism that could capture the common conception of morality, according to which morality is objective and categorically normative, while circumventing the most daunting problems associated with moral realism. I first define moral realism and discuss its problematic features. I claim that while non-naturalism is better suited than naturalism to account for the categorical normative force of morality, stance-independent versions of non-naturalism have serious metaphysical and epistemic problems. A possible alternative is to adopt a non-ontological non-naturalism and complement it with a Kantian and constitutivist origins story. I then explore evolutionary debunking arguments, which pose the most pressing epistemic problem that all non-naturalist accounts of moral objectivity face, and argue that they are not strong enough to undermine non-naturalism. Finally, I discuss different versions of constructivism and explain why transcendental constitutivism, which is a stance-dependent non-naturalist view, has the potential to account for objectivity and categorical normativity while avoiding the problems associated with realism. My claim is that transcendental constitutivism is a neglected alternative to moral realism.

KANTIAN CONSTITUTIVISM AS AN ALTERNATIVE TO MORAL REALISM

A DISSERTATION

SUBMITTED ON THE THIRD DAY OF MAY 2022

TO THE DEPARTMENT OF PHILOSOPHY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

OF THE SCHOOL OF LIBERAL ARTS

OF TULANE UNIVERSITY

FOR THE DEGREE

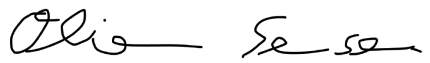
OF

DOCTOR OF PHILOSOPHY

BY



Caner Turan

APPROVED: 

Oliver Sensen, Ph.D.
Director



Alison Denham, Ph.D.



Bruce Brower, Ph.D.

©Copyright by Caner Turan, 2022
All rights reserved

ACKNOWLEDGEMENTS

I would like to thank all the members of my committee, Oliver Sensen, Alison Denham, and Bruce Brower, for very helpful feedback, suggestions, and guidance. Without their advice the dissertation could not have taken its present shape, and my future development of these ideas would be impaired. I would also like to thank my family for their continued support.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	ii
INTRODUCTION.....	1
CHAPTER 1: MORAL REALISM.....	10
1.1 The Importance.....	11
1.2 The Definition.....	13
1.3 Forms of Moral Realism.....	18
1.3.1 Non-Naturalism vs. Naturalism.....	18
1.3.2 Supervenience and Analytical Naturalism.....	21
1.3.3 Open Question Argument and Synthetic Naturalism.....	24
1.4 Problems.....	28
1.4.1 Moral Twin Earth.....	29
1.4.2 Normativity.....	32
1.4.3 Supervenience.....	38
1.4.4 Moral Epistemology.....	55
CHAPTER 2: EVOLUTIONARY DEBUNKING ARGUMENTS.....	79
2.1 How to Respond to Evolutionary Debunking Arguments.....	80
2.1.1 Evolutionary Challenge.....	81
2.1.1.1 Structure of Evolutionary Debunking Arguments.....	81
2.1.1.2 Street.....	83

2.1.1.3 Joyce.....	85
2.1.2 Ambition and Strength.....	87
2.1.2.1 The Inverse Rule of Debunking.....	88
2.1.2.2 Self-Refutation Argument.....	89
2.1.2.3 Ambition and Inductive Strength.....	93
2.1.3 Logical Property vs. Dialectical Property.....	94
2.1.4 Conclusion.....	97
2.2 Greene and the Weakness of the Empirical Premise.....	97
2.2.1 Greene’s Argument.....	101
2.2.2 Personal Force.....	104
2.2.3 Explaining Different Reactions.....	105
2.2.4 Types of Moral Intuitions and Autonomous Moral Reasoning.....	109
2.2.4.1 Types of Moral Intuitions.....	109
2.2.4.2 Autonomous Moral Reasoning.....	113
2.2.5 Formal Intuitions.....	123
2.2.6 How to Acquire Theoretical Intuitions.....	128
2.2.6.1 Two Systems.....	129
2.2.6.2 Internalization of Reasons and Cultural Evolution.....	132
2.2.6.3 Dawkin’s Meme Theory.....	138
2.2.6.4 Our Reactions to Moral Cases.....	140
2.2.7 Conclusion.....	142
CHAPTER 3: KANTIAN CONSTITUTIVISM.....	144
3.1 The Constructivist Project.....	149

3.2 Idealized Stance Constructivism.....	153
3.2.1 Idealized Stance Constructivism and Objectivity.....	155
3.2.2 Normative Restrictedness.....	159
3.2.3 Euthyphro Dilemma.....	161
3.2.4 Conclusion.....	164
3.3 Constitutivism.....	165
3.3.1 Velleman.....	167
3.3.2 Street.....	171
3.3.3 The Basic Constitutivist Strategy and Inescapability.....	173
3.3.4 Humean vs. Kantian Constitutivism.....	175
3.3.5 Problems with Humean Constitutivism.....	180
3.4 Kantian Constitutivism.....	186
3.4.1 Kant against Realism.....	192
3.4.2 Korsgaard.....	195
3.4.3 Problems with Korsgaard.....	201
3.4.4 Transcendental Constitutivism.....	209
3.4.4.1 First Step.....	212
3.4.4.2 Second Step.....	219
3.4.4.3 Third Step.....	223
3.4.5 Parfitian Moral Ontology.....	224
3.4.6 Moral Intuitions.....	232
3.5 Objections to Transcendental Constitutivism.....	237
3.5.1 Appeal to a Noumenal Realm.....	237

3.5.2 The Bootstrapping Objection.....	250
3.5.3 Evolutionary Challenge Revisited.....	257
REFERENCES.....	266

INTRODUCTION

“Truths do not have to exist, or be real, in an ontological sense. Truths need only be true.” – Derek Parfit, On What Matters¹

“[Morality] is to be firm even though there is nothing in heaven or on earth from which it depends or on which it is based.” – Immanuel Kant, Groundwork of the Metaphysics of Morals²

We often judge certain practices to be right, wrong, good, or bad. For example, we think that unjustified murder is bad. We think that helping the poor is the right thing to do. We make these judgments with a high degree of certitude. It is therefore generally believed that we are committed to moral objectivity, at least in our intuitive starting points. Even people who think that moral truth is relative or who think that moral judgments are merely expressions of our emotions or other conative states accept the compelling *appearance* of objectivity in morality. This is our common conception of morality.

Moral realism captures the common conception of morality. The strongest appeal of moral realism, therefore, is its adherence to objective moral truths. According to realism, if there are to be objective moral truths, moral properties and facts must be *stance-independent*. That is, morality must exist ‘out there’ (or must be tied to human rationality as an absolute inner value), as part of the fabric of the universe *independently* of any

¹ Parfit 2011b, 21.

² GMS 4:425.

human perspective. Just as the basic facts of logic or physics do not depend in any way on human endorsement, the realist asserts, so the facts of morality are *discovered* rather than created by humans.

The stance-independence condition of both naturalist and non-naturalist versions of realism is subject to serious metaphysical and epistemic objections, which makes it difficult to accept such a view. If moral properties and facts are independent of us, then we are alienated from moral truths in metaphysically and epistemically problematic ways. The aim of this dissertation is to explore alternatives to moral realism that could capture the common conception of morality, according to which morality is objective and categorically normative, while circumventing the most daunting problems associated with moral realism.

In the first chapter, I define moral realism and discuss the problems associated with it. First, I adopt a metaphysical rather than a semantic definition of moral realism to distinguish it from other views that allow for moral truth, and to reveal its most distinctive and appealing feature, namely stance-independence.

Second, I show that both naturalistic and non-naturalistic conceptions of stance-independence are hard to defend. It seems quite difficult to retain the objectivity or genuine normative force of morality within a naturalistic framework. Merely listing the natural features of actions does not tell us *why* those features meet or fail to meet the standards of goodness or badness for human action. The naturalist must show that there are naturalistically characterizable action-guiding standards that are also *categorically* reason-giving, which seems to be an impossible task.

While non-naturalism is better suited than naturalism to account for the categorical normative force of morality, non-naturalism seems to be unable to explain the supervenience of the moral on the natural. I argue against Shafer-Landau's (2003) 'exhaustive constitution thesis' and FitzPatrick's (2008) 'dual-aspect' solution. I claim that the former collapses into naturalism, while the latter offers a bloated ontology and fails to capture the *one-way* necessary dependence of moral properties on natural ones. Furthermore, I argue that evolutionary debunking arguments pose the most pressing epistemic problem that non-naturalist accounts of objectivity face.

In the second chapter, I discuss evolutionary debunking arguments (EDAs). EDAs generally attack moral realism (or stance-independent accounts of moral objectivity in general), but evolutionary explanations of our moral intuitions and beliefs may pose a problem for *all* kinds of non-naturalism. Kantian constitutivism, which I discuss in the third chapter, is also a form of non-naturalism (despite being a *stance-dependent* account) and it is therefore another target of EDAs. So, it is important to address this epistemic issue.

The idea that moral intuitions and beliefs are determined exclusively by evolutionary processes seems to be a fundamental threat to assumptions that underlie this dissertation. There are two assumptions that underlie my dissertation. (I explain in detail why I have these assumptions in the first chapter.) These two main assumptions constitute the starting point of my inquiry and inform my further investigations: (1) morality, if it exists, is *objective*, and (2) morality, if it exists, is *categorically normative*. Perhaps we think morality is objective and categorically normative not because morality really is so but because having such beliefs or intuitions makes us good social cooperators. Richard

Joyce (2001, 141–8) offers a compelling story of how evolutionary forces could have shaped our sense of ‘moral ought’ and our capacity for normative guidance. The idea of an evolutionarily acquired innate moral sense seems to be at odds with all non-naturalist accounts of moral objectivity and categorical normativity because if our moral sense is implanted in us by evolutionary processes, then morality cannot be necessary and universal: we could have had a different moral sense under different circumstances.

EDAs, *prima facie*, pose a serious epistemic threat to non-naturalist accounts of moral objectivity. However, (a) they are not strong enough to undermine the position (realism or constitutivism), and (b) the alternative account that I discuss in the third chapter (Kantian constitutivism) is compatible with the idea that our sense of moral objectivity and categorical normativity have increased our reproductive success. I discuss (a) in the second chapter, and (b) in the third chapter.

To show (a), I claim that the ambition of an EDA affects the argument’s *empirical* premise: the more set of beliefs an EDA calls into question the harder it becomes to provide a complete evolutionary origins story. This is because our theoretical and formal moral intuitions are immune to direct evolutionary influence. While theoretical intuitions seem to be the product of reflection on evolved psychological dispositions and of the process of cultural evolution, formal intuitions seem to be the product of what is entailed by the nature of moral concepts and what is entailed by the constitutive features of rationality.

I use the talk of moral intuitions and use Michael Huemer’s (2008) categorization of moral intuitions to explain why evolutionary debunking arguments are not strong enough to undermine non-naturalist accounts of moral objectivity and categorical normativity.

There are four types of moral intuitions: (i) concrete intuitions; (ii) mid-level intuitions; (iii) theoretical intuitions; and (iv) formal intuitions.

This does not mean that I adopt rational intuitionism like Huemer. On the contrary, I argue against it in the first chapter. In the third chapter, I claim that formal moral intuitions could be explained by how our reason necessarily functions (or what is entailed by the constitutive features of rationality). And theoretical intuitions arise partly from the application of *formal* intuitions to the *matter* provided by the details of one's emotional and social context. So, this account is compatible with the version of Kantian constitutivism I support.

Theoretical intuitions are the product of systematic reflection on our reactive attitudes towards particular cases (concrete and mid-level intuitions). After reflecting systematically on our concrete and mid-level intuitions, we reach generalizations or abstract moral theories (e.g., It is wrong to use people as mere means).

Formal intuitions are the product of what is entailed by moral concepts, our attitude of valuing, and constitutive features of rationality. The following are examples of formal intuitions:

- (1) If x is better than y and y is better than z , then x is better than z .
- (2) If x is a reason for y , it is not the case that that x is not a reason for y .
- (3) If only facts of kind x are reasons for y , and z is not of kind x , then z is not a reason for y .
- (4) If it is wrong to do x , and it is wrong to do y , then it is wrong to do both x and y .
- (5) If two states of affairs, x and y , are so related that y can be produced by adding something valuable to x , without creating anything bad, lowering the value of anything in x , or removing anything of value from x , then y is better than x .

I then describe one possible way of turning/converting our generalizations or abstract moral theories into intuitions. Instead of trying constantly to keep our pre-reflective

intuitions in check, we have created a set of shared values, *some* of which are the product of autonomous reasoning, and, more importantly, we have *internalized* them: they have become second nature to us.

In the third chapter, I discuss a neglected alternative to moral realism, namely Kantian constitutivism. The main problem of realism is the conviction that moral properties exist ‘out there’ (or must be tied to human rationality as an absolute inner value), i.e., they are part of the fabric of a stance-independent reality. Modeling moral ontology on scientific ontology not only poses seemingly unsolvable metaphysical problems but also runs the risk of conflating the kind of objectivity possessed by empirical facts with the kind of objectivity possessed by moral facts. While empirical facts are about ‘what *is*,’ moral facts are about ‘what *ought* to be.’ Given this fundamental difference, it could make more sense to claim that moral facts, unlike empirical ones, exist in a *non-ontological* way like mathematical and logical facts. This Parfitian view, however, is as mysterious as other non-naturalistic views if we do not complement it with a plausible *origins* story.

I evaluate different constructivist views to see whether they can provide us with such a story. Subjectivism, relativism, ideal observer theories, procedural constructivism, and Humean constitutivism fail to accomplish the task as they all weaken the standard of objectivity by basing morality on desires, preferences, intersubjective agreements, or contingent evaluative starting points. This is not to say that these forms of constructivism are *wrong*. I do not reach such a dogmatic conclusion. My claim is only that Kantian constitutivism has the potential to account for objectivity and categorical normativity while avoiding the problems associated with realism. And while it may have its own problems, it is a neglected alternative. I do not claim anything more than this.

Korsgaard's (1996, 2009) Kantian constitutivism seems to fare better for our purposes, since it takes morality to be necessary and universal yet still dependent on human reason. However, Korsgaard ultimately reduces moral normativity to the normativity of consistency and her account either collapses into realism or it can at most justify *subjectively* universal principles. The problem seems to be that Korsgaard takes the *executive* function of practical reason (*Willkür*) as the source of morality. Another, and possibly more plausible, way to defend Kantian constitutivism is to focus on the *legislative* function of practical reason (*Wille*).

On this alternative view, namely transcendental constitutivism, the moral law is not an agency-enabling principle that provides psychological unity, but it is constitutive of how *pure* reason operates. That is, the moral law does not constitute one's *choices*, but it is a necessary law that guides one's consciousness in thinking about practical matters, similar to how laws of logic govern theoretical thinking. Moral value is conceived of as an operating principle of how our reason *functions* rather than as an *entity* or *property* that shows up on an ontological radar screen. This view captures the objectivity and categorical normativity of morality unlike other forms of constructivism, and it circumvents the problems associated with the ontological characterization of moral value and the perceptual characterization of moral knowledge.

I explain why the account of moral intuitions I described in the second chapter is compatible with transcendental constitutivism. I claim that formal moral intuitions could be explained by how our reason necessarily functions (or what is entailed by the constitutive features of rationality). And theoretical intuitions could arise partly from the

application of *formal* intuitions to the *matter* provided by the details of one's emotional and social context.

I then address possible objections to transcendental constitutivism. First, the view I defend does not appeal to a noumenal realm due to the distinction between independence from *causation* in nature and independence from *existence* in nature. The fact that pure reason gives the moral law spontaneously does not entail that pure reason is ontologically independent of the natural world. Rather, we can conceive of pure reason or freedom (*Wille*) as residing in nature as an emergent, unalterable structure of thinking.

Admittedly, the view in question is a non-trivial view about how reason functions. And we have no conclusive proof that we are really free in the way Kant describes it.

However, it is possible to find allies of this view, which place the function of reason at the center, such as Chomskyan linguistics, Mikhail's universal moral grammar, functionalism in the philosophy of mind, and Jerry Fodor's modularity of mind.

Second, the objection that no substantive moral content can be extracted from Kant's Categorical Imperative (CI) is hasty. Admittedly, a purely formal law, considered in and of itself, is empty. Kant himself seems to agree with this. However, if our reason draws inferences from *empirically identifiable* universal human ends to determine specific moral content, and makes the necessary means to those universal ends *binding* through the CI, the objection could fail.

Third, evolutionary debunking arguments (EDAs), *prima facie*, pose a serious epistemic threat to the non-naturalist accounts of moral objectivity; however, the proposed account is compatible with the idea that our sense of moral objectivity and categorical normativity have increased our survival and reproductive success.

I defend the idea that our capacity to be spontaneous, or to be free from the influence of nature, namely pure reason or freedom, is the source of morality, and that it could have been evolved in nature. However, our sense of moral ought, the intuition that we *can* and *should* choose the morally right thing despite our strongest desires, follows *necessarily* from freedom (*Wille*) itself rather than from the forces of natural selection. Having this moral sense could have had a positive effect on our biological fitness as Joyce (2001, 2006) says, but no evolutionary history is needed to account for the relation between reason and the CI. This is because the CI arises *a priori* from reason (or freedom). That is, the CI is the *byproduct* rather than the direct product of evolution. If we could create a free being, a being with a mind that has a certain level of complexity, *out of nothing*, that being would be under the CI and would have a sense of moral ought.

This is against evolutionary explanations such as Joyce's, according to which internalization of a moral sense requires evolutionary history. However, the Kantian view I defend does not entail that individuals can create morality (as we know it) or moral behavior in isolation from each other. On the contrary, morality is realizable for creatures like us through the development of a social community. That is, if a being possesses a reason that is free, she is immediately under the moral law. The content of the moral law "does not alter;" it is independent of our evolutionary history. Nevertheless, to be able to derive specific moral rules and to reach moral judgments, we need evolutionary history. If we did not live together in groups, we would not have to justify ourselves to each other. We would not think about moral reasons for actions. We would only be under hypothetical imperatives (e.g., do X in order to avoid pain).

CHAPTER 1: MORAL REALISM

My aim in this chapter is to set the stage for the assessments in the following chapters by providing a definition of moral realism and by discussing the metaphysical and epistemological problems generally associated with different versions of it. First, I will talk about the importance of moral realism (1.1) and define the view (1.2). I will then distinguish between different versions of moral realism (1.3) and discuss the objections directed at each version (1.4).

Obviously, I cannot provide a detailed discussion of moral realism in a dissertation chapter. Indeed, each objection I mention deserves more careful and lengthy consideration than I can give here. My aim, however, is not to solve the problems of realism within the realist framework. Rather, my aim is to offer an alternative to moral realism that can capture what is most attractive about moral realism while at the same time circumventing the most daunting problems associated with it. It is therefore essential to think about what makes realism appealing and what specifically makes it hard to accept such a view. This is what I do in this chapter. I do not, however, claim to be comprehensive in my discussion.

The results of my discussion can be summarized in six points: (1) An adequate definition of moral realism should reveal what most people find attractive about such a view, namely the existence of objective moral truths; (2) The basic version of the Open Question Argument is effective against analytic naturalism but it cannot refute

metaphysical naturalism; (3) However, the Normativity Objection, which is a contemporary extension of the Open Question Argument, is a serious problem for metaphysical naturalism; (4) Explaining supervenience is a big problem for non-naturalist metaphysics; (5) Evolutionary debunking arguments pose the most pressing epistemic problem that non-naturalist realism faces; (6) If it is possible to talk about *stance-dependent* objective moral truths with genuine normative force, then it may also be possible to offer (i) a parsimonious moral ontology that is compatible with moral supervenience and (ii) an epistemology that is compatible with distorting evolutionary influence on our moral beliefs.

1.1 The Importance

We judge certain practices or actions to be right or wrong. For example, we think that torturing for fun is wrong; we regard helping the poor as good, and so on. Even though we make such moral judgments frequently, we do not think much about *what* makes actions right or wrong (unless we are moral philosophers). A general moral principle could do the job. For instance, what makes murder wrong could be that it is against the utilitarian principle of maximizing happiness. Or murder could be wrong because it is against the deontological principle that one should never treat anyone merely as a means. Alternatively, a virtue ethicist could think that murder is wrong because a person who has developed virtuous character traits would not commit murder. The investigation of the general principles about what is right or wrong is called ‘normative ethics.’

Moral philosophers also examine specific controversial issues such as euthanasia, death penalty, famine relief, animal rights, protection of the environment, and so on. In

doing so, they usually apply what they take to be the right moral principle to particular situations and try to solve practical problems. For example, one could claim that euthanasia is permissible because it decreases the amount of misery in the world. This level of ethical inquiry is called ‘applied ethics.’

Even if we could find a moral principle applicable to all situations, there would still be an important question to be answered, namely, what is the *nature* of this principle? Does the principle hold independently of our desires and inclinations, or does it merely express our preferences? These are *metaethical* questions. Metaethics asks questions about the metaphysical nature of moral properties, the possibility of moral knowledge, the meaning of moral terms, and the proper moral motivation, among others. Simply put, one is engaged in a metaethical inquiry when one asks questions about the nature of morality and the meaning of moral judgments.

Moral realism is a theory in metaethics. It claims that there are *objective* moral truths. That is, according to moral realism, there are moral statements that are true independent of our attitudes toward them. For instance, to say that “genocide is wrong” is an objective moral truth is to say that genocide is wrong regardless of what anyone thinks or feels about it. If moral realism is the correct metaethical theory, then moral truth cannot be determined by individual or group preferences, desires, conventions, or agreements. Just as the basic facts of logic or physics do not depend in any way on human endorsement, the realist asserts, so the facts of morality are *discovered* rather than created by humans.

One of the appeals of moral realism is that it captures our commitment to morality’s objectivity. Whenever we engage in a moral discussion in our daily lives, and specifically when we discuss about moral issues we deeply care about (e.g., gender-based

discrimination, protection of the environment, women's rights, racism, and so on), we think that our statements about these issues would still have been true even if we had different beliefs and practices. For example, you may disagree with someone who supports female genital mutilation (FGM). Although you may grant that their support is rooted in their culture or their individual preferences, you will *believe*, deep down, that there is something genuinely *wrong* with FGM regardless of what anyone or any culture thinks about it. Thus, you will think that they are missing an *objective* moral truth, perhaps due to their wrong factual belief that excision is beneficial to society. The phenomenology of moral discussion and disagreement indicates that ordinary moral discourse *aspires* to objectivity.¹ Arguably, we are all moral realists at least in our intuitive starting points, and therefore it is important to reflect on the plausibility of such a view.²

1.2 The Definition

It is also important to distinguish moral realism from other metaethical views. The FGM example may help in defining the view. There are *three* features that characterize your disagreement about FGM. First, you *believe* that FGM is wrong. In other words, you try

¹ Cf. Enoch 2014

² Expressivists reject that we are committed to morality's objectivity. They argue that moral statements do not attempt to describe the way the world is, but rather they are used to express certain motivational states that are different from beliefs, such as approval, preference, commitment to a norm or a plan, and so on (Blackburn 1984, 1998; Gibbard 1990, 2003). If moral statements do not express beliefs, then they cannot be true or false. Hence, no aspiration to objectivity. Nevertheless, even expressivists acknowledge that moral statements act exactly like proposition-expressing statements. So, they accept the compelling *appearance* of objectivity in morality and try to account for it in their increasingly complicated theories.

to *describe* some aspect of the world. Second, one of you is *correct* about the moral status of FGM. However, you cannot both be correct. Only one of you has a true moral belief and gets the moral facts right. Third, the fact that determines who is right is independent of people's attitudes toward FGM.

According to a popular definition given by Geoffrey Sayre-McCord, adopting the following two theses makes one a moral realist: "(1) the claims in question, when literally construed, are literally true or false (cognitivism), and (2) some are literally true." (Sayre-McCord 1988, 5) This definition attempts to distinguish realist positions from anti-realist ones in terms of what we *mean* when we sincerely utter moral sentences. Thus, it is a *semantic* definition. If Sayre-McCord's definition is correct, then the following two features of moral judgments are individually necessary and jointly sufficient to characterize moral realism: (i) moral judgments are intended to *describe* how the world is by attributing moral properties to things (Cognitivist Claim), and (ii) some moral judgments are true in virtue of providing an *accurate* description of the way the world is (Success Claim). That is, *any* metaethical view that embraces a 'cognitivist success theory' *is* realist under this definition.

It seems that there is something incomplete about Sayre-McCord's definition, especially if we think about it in connection with the disagreement example. His definition only captures the *first* feature of moral disagreement, namely, that moral judgments express beliefs. The definition also allows the existence of moral truth, but it does not specify *what* makes moral beliefs true. Therefore, accepting it would force us to classify *any* metaethical view that allows for moral truth as realist. For example, Railton's (1986) and Smith's (1994) naturalistic theories that define moral truth in terms

of desires or psychological responses of a fully rational and informed observer would be realist under the semantic definition.³ On top of that, according to the crudest versions of subjectivism and relativism, it could be *true* for me (subjectivism) or my culture (relativism) that FGM is wrong, whereas it could equally be *true* for someone else or some other culture that FGM is right. These views would also count as realist if we were to accept Sayre-McCord's definition.

Furthermore, the semantic definition fails to distinguish moral realism from sophisticated forms of expressivism, according to which moral judgments express attitudes of approval, disapproval, endorsement, criticism, and the like, toward non-moral circumstances. Most of the sophisticated forms of expressivism allow for talk of moral truth by providing a deflationary account of truth.⁴ For example, if the term 'true' does not represent any property but is simply used to *endorse* or *affirm* the content of a moral sentence, then expressivists can ascribe truth to moral judgments without committing themselves to independent moral facts as truth-makers of such judgments. This means that advocates of some forms of expressivism may be happy to support a cognitivist success theory, and thus they may also be called moral realists if we stick to Sayre-McCord's definition.

The problem is that merely allowing the existence of moral truth does not make one a moral realist. To call each view that allows for moral truth realist would surely create terminological confusion.⁵ What is more, to classify subjectivist, relativist, ideal

³ See 1.4.2 for more on Railton and Smith.

⁴ For example, Blackburn 1984; Gibbard 2003.

⁵ Moreover, the term 'realism' would lose its philosophical significance since many incompatible theories would count as realist.

observer, and expressivist theories as different forms of realism would disguise what most people think is appealing about moral realism, namely, that moral properties exist ‘out there’ (or they are tied to human rationality as an absolute inner value) as part of the fabric of the universe independently of *any* human perspective. This is the third feature of the phenomenology of moral disagreement, namely, that people’s contingent attitudes cannot affect the moral status of actions. If you sincerely believe that FGM is immoral, you think it is wrong no matter what people think or feel about it. If you think racism is bad, you think it is bad not because we *want* or *need* it to be bad but because racism is inherently or ‘really’ bad. There is something odd and confused about the claim that morality is ‘real’ if you also think that moral truth is created by a *contingent* human perspective. This is because we are committed to morality’s objectivity in our intuitive starting point, and if something is *real* it *is* what we *think* it is. If morality is real, it is objective. If morality is not real, it is not objective. Thus, a ‘real’ moral truth should be an ‘objective’ one.

Although Sayre-McCord’s definition has been supported by some metaethicists,⁶ there has been a trend in the recent metaethical literature to work with a *metaphysical* definition of realism instead of a semantic one.⁷ This trend is not surprising considering the implications of the semantic definition. It is called the ‘metaphysical’ definition because it is marked by a metaphysical commitment to objective moral properties and facts. Metaphysical definition attributes a third feature to moral realism: (iii) moral judgments are made true *independently* of our intentional attitudes and our conceptual

⁶ For example, Cuneo 2007, 45–9.

⁷ Shafer-Landau 2003; Dreier 2004; Huemer 2005; FitzPatrick 2008; Miller 2009.

schemes. We can call this feature the ‘Objectivity Claim,’ since it successfully captures morality’s aspiration to objectivity. Metaphysical definition of moral realism is usually associated with Shafer-Landau’s *stance-independence* condition: “there are moral truths that obtain independently of any preferred perspective, in the sense that the moral standards that fix the moral facts are not made true by virtue of their ratification from within any given actual or hypothetical perspective.” (Shafer-Landau 2003, 15) In other words, on Shafer-Landau’s definition, a view that does not refer to *objective* or *stance-independent* moral properties and facts in explaining the rightness or wrongness of actions is anti-realist, regardless of its take on the semantics of moral discourse. I adopt Shafer-Landau’s metaphysical definition of moral realism because, as I mentioned, I think that morality is real only if it is objective.

There are two things to keep in mind about the stance-independence condition. First, the condition entails that moral facts are independent of what we choose or want, regardless of what *kind* of moral properties they instantiate. Realists could characterize moral properties such as goodness or rightness as natural *or* non-natural properties. However, such a difference in characterization does not change the fact that they regard moral facts as objective.

Second, the stance-independence condition does *not* entail that moral facts are independent of the *existence* of human beings. For example, it could turn out that only human (or humanlike) beings have a value property that is attached to their rational nature.⁸ That would make the existence of moral facts dependent on the existence of human (or rational) beings. However, that would not entail that we *create* or confer value

⁸ Langton (2007) argues for such a view.

on ourselves by actual or hypothetical choices. Rather, we would have that (absolute inner) value property prior to any choice we make. This means the value property in question would not be constructed but *discovered* in successful ethical inquiry.

It is important to note that, according to the realist, moral *standards* or *principles* would still exist even if there were no human beings. For instance, Shafer-Landau characterizes moral principles as conditionals of the form “If X instantiates the (non-moral) property P, it also instantiates the moral property M,” while he characterizes moral *facts* as “instantiations of moral properties” that make our moral judgments true (Shafer-Landau 2003, 83n36, 268n127). Absent all human beings and their mental states, there would be nothing to instantiate the property P; hence, no moral facts. However, the conditional form of moral principles would hold even in that case.

1.3 Forms of Moral Realism

1.3.1 Non-naturalism vs. Naturalism

The distinguishing feature of moral realism is its adherence to objective moral truths. That is, all forms of moral realism are committed to the stance-independence of moral properties and facts. Therefore, if the Objectivity Claim or the stance-independence condition is philosophically problematic, then this is a common problem for *all* forms of moral realism.

However, there are also specific metaphysical and epistemic objections raised against specific types of moral realism. Different ways of characterizing moral properties or different takes on how to attain moral knowledge bring up different metaphysical and epistemic issues. So, it is important to distinguish between different types of moral

realism and to reflect on the problems generally associated with each. My aim is not to provide an extensive assessment of all types of moral realism, but rather to give reasons why I find certain objections more serious than others and, relatedly, why I focus on certain issues rather than others throughout my dissertation.

What kinds of stance-independent facts could make moral statements such as “X is good” true? There are two popular options: (1) X has a *non-natural* property, or (2) X has a *natural* property such as health, pleasure, or happiness. The former view is called ‘moral non-naturalism,’ while latter is called ‘moral naturalism.’⁹ Although both views argue for the existence of objective moral facts, they differ in how they characterize the nature of moral properties. Moral naturalism identifies moral properties (and facts) with natural properties (and facts), as the name suggests, while moral non-naturalism rejects the equation of the moral with the natural.

There is no uncontroversial answer to the question of what constitutes the *naturalness* of a property; however, an intuitively plausible way of defining natural properties is to view them as properties that are studied by natural sciences.¹⁰ In other words, natural properties are those that are constituted by arrangements of elementary particles in physics. It also makes some sense to define natural properties in terms of their causal efficacy. That is, to be a natural property is to have causal effects on the world.¹¹ Natural properties can also be characterized semantically or epistemologically. From a linguistic

⁹ To name a few examples, Moore (1903), Shafer-Landau (2003), Huemer (2005), FitzPatrick (2008), Enoch (2011), and Parfit (2011) are moral non-naturalists, while Sturgeon (1985), Brink (1986), Boyd (1988), Jackson (1998), and Finlay (2014) are moral naturalists.

¹⁰ Cf. Moore 1903, 55; Shafer-Landau 2003, 59.

¹¹ Copp 2003, 183–4; Sturgeon 2006, 100–1; Bedke 2009.

perspective, natural properties are properties that we can describe using non-evaluative terms such as water, flower, or silver.¹² And from an epistemological perspective, natural properties are those that can only be investigated through empirical methods.¹³ While moral naturalism sees morality as being *continuous* with the natural sciences both in its subject matter and its methodology, moral non-naturalism views morality as being *autonomous* from the natural sciences.

This autonomy can be understood metaphysically or epistemologically. First, moral properties could be metaphysically different from natural ones. Goodness could be a *sui generis* and irreducible non-natural property. This would mean that an action's being good is not just a matter of its having a real property picked out by the word 'good,' but it is also a matter of its having a *sui generis evaluative* property which is *irreducible* to any natural property. This is probably the most distinctive feature of G. E. Moore's (1903) non-naturalism,¹⁴ adopted also by contemporary moral realists such as FitzPatrick (2008). Alternatively, one can claim that moral (or normative) properties *exist*, but they exist in a *non-ontological* sense like mathematical or logical properties (Parfit 2011b, chapter 31). There are true mathematical or logical claims, but they are true not because they accurately describe a spatio-temporal or non-spatio-temporal part of reality. Mathematical or logical truths are discovered through first-order mathematical or logical reasoning rather than a second-order metaphysical investigations into what mathematical or logical properties exist. Likewise, the idea goes, moral truths are discovered through

¹² Jackson 1998, 7.

¹³ Little 1994, 226; Copp 2003, 181; Shafer-Landau 2003, 55.

¹⁴ Cf. Pigden 1993, 421–2.

first-order moral reasoning rather than metaphysical investigation. That is, we discover moral truths through the employment of internal, domain-specific methods and standards.

Second, the non-naturalist autonomy claim can be understood epistemologically. For example, Shafer-Landau (2003) defines natural properties as those that are discoverable only empirically, but he then asserts that moral facts are knowable only by a special faculty of intuition. This commitment makes him a non-naturalist only in an epistemic sense because he also argues that moral properties are *exhaustively constituted* or realized by some arrangement of a series of natural properties. According to Shafer-Landau (2003, 76–8), even though no moral property is identical to any natural property or properties, each instantiation of a moral property is exhaustively constituted by natural properties. So, a realist view can be non-naturalist about moral epistemology while at the same time embracing a naturalist but non-reductive moral ontology.

There are then two main forms of realism: non-naturalism and naturalism. Just as there are different versions of non-naturalism, there are different versions of naturalism.

1.3.2 Supervenience and Analytic Naturalism

Although all forms of naturalism share the commitment to the metaphysical idea that moral facts are natural facts, we can distinguish between *two* main types of naturalism with different epistemic and semantic commitments: (reductive) analytic and (non-reductive) synthetic naturalism.¹⁵

Analytic naturalism is motivated by the problem of explaining *supervenience* of the moral on the natural in non-reductive terms. Non-naturalism conceives of moral and

¹⁵ Synthetic naturalism is also known as Cornell realism.

natural properties as distinct existences that are not reducible to each other. But if this is true, then it is quite difficult, if not impossible, to explain the necessary dependence of the moral properties on the natural ones. Imagine a possible world W that is exactly like our world in all of its naturalistic or descriptive features. Our world and W would be identical in their moral properties as well, since we do not have any reason to claim that the moral status of, say, murder would differ in two naturalistically identical worlds. However, descriptive differences do not entail moral differences. For example, grass could be purple in W , but all else being equal, murder would still be wrong. Now, if there can be no change in moral properties without a change in natural properties, that is, if moral facts are entirely fixed by natural ones, then moral properties must be *reducible* to natural properties. It seems in principle impossible to explain supervenience if you hold that the moral cannot be reduced to the natural.¹⁶

Supervenience is one of the least controversial theses in metaethics,¹⁷ and it seems to entail that moral and natural properties are necessarily coextensive. In other words, it seems impossible for moral and natural properties not to coincide. If this is the case, then we must be talking about a single property instead of two distinct properties. And considering the fact that moral properties are entirely determined by natural ones, the thought goes, it must be that moral properties *just are* natural properties.¹⁸ And if so, moral and descriptive statements must refer to the same facts and properties.

¹⁶ For more on supervenience, see 1.4.3. See also Mackie 1977, 41; Cuneo 2007; Enoch 2011; McPherson 2012.

¹⁷ Cf. Rosen 2020.

¹⁸ Cf. Jackson 1998, 124. For an objection to this line of thinking, see Parfit 2011b, 296–7.

The defining feature of reductive analytic naturalism is the semantic thesis that moral and natural terms are equivalent in meaning. The assumption seems to be that moral and natural terms pick out the same property *only if* they are synonymous. If moral and natural terms are synonymous, then we can explain truths about what is morally good in terms of natural facts. For example, suppose you think that an action is morally good if it promotes health. To determine whether an action is good, you just need to empirically investigate whether it promotes health. Moral facts, such as “X is good,” are reducible to natural (or descriptive) facts, such as “X promotes health,” if analytic naturalism gets things right.¹⁹

This semantic commitment comes with epistemic implications. Recall the epistemic definition of moral properties as properties that can only be known empirically (or *a posteriori*). Analytic naturalism is at odds with that definition, for the synonymy between moral and natural terms can in principle be revealed by conceptual analysis. This means that moral standards such as “X is moral iff X promotes health” can be known *a priori* rather than empirically. Nevertheless, analytic naturalists think that most of the content of morality is discoverable empirically. It might be an analytic truth that “Exercise is morally good” just means “Exercise promotes health.” However, whether exercise promotes health is a matter of empirical investigation, i.e., it is knowable only *a posteriori*. The fact that exercise improves the strength of heart and lungs, for example, can only be known empirically. Thus, according to the analytic naturalist, only *some* moral claims are knowable *a priori*. Although the inclusion of *a priori* moral knowledge is not *fully* compatible with the naturalistic aim of applying scientific methodology to

¹⁹ Frank Jackson (1998) and Stephen Finlay (2014) are renowned analytic naturalists.

ethical inquiry, analytic naturalism enjoys a *prima facie* advantage of being able to explain the supervenience of the moral on the natural.

1.3.3 Open Question Argument and Synthetic Naturalism

There is an important implication of the synonymy between two terms: identity questions of the form “is it true that X is Y?” are meaningless if they are asked about analytically equivalent terms. Think about synonymous terms such as ‘bachelor’ and ‘unmarried.’ It is a trivial truth that “if S is bachelor, then S is unmarried.” Since this is a conceptual truth, we cannot make sense of someone accepting that S is bachelor but wondering whether S really is unmarried. “I know that S is bachelor, but is S unmarried?” is a closed question. It is meaningless to ask such questions. This suggests that asking moral-natural identity questions must similarly be meaningless or inherently confused given the synonymy between moral and natural terms. But such questions do not strike us as meaningless or inherently confused: they are *open* questions. This is Moore’s (1903) Open Question Argument against naturalism.

According to the Open Question Argument, for any naturalistic description N of a moral term M, it will always be possible for a competent user of moral terms to acknowledge that something has the proposed natural property N but still ask whether it is really M. For example, suppose someone asserts that ‘being good’ is synonymous with ‘being pleasurable.’ The question “Is it good to be pleasurable?” must mean the same as the question “Is it good to be good?” due to the analytic equivalence between these terms. If so, then “Is it good to be pleasurable?” must be a meaningless question. But it is not. It

is an open question whether being pleasurable is good. Therefore, ‘good’ and ‘pleasurable’ (or any natural term) cannot be synonymous.

Although the Open Question Argument cannot deliver a knockout blow to analytic naturalism, it is still effective, and it is effective *only* against analytic naturalism. If you are an analytic naturalist, I think the most plausible response you can have is to claim that the openness of moral-natural identity questions does not entail that moral predicates are unanalyzable. If goodness is analyzable in non-normative terms, then every analysis will be false but the correct one. It is not surprising that the questions of the form “is X, which is N, good?” will feel open because it is possible that we have not yet come up with the correct analysis.²⁰ This response is not satisfactory because we seem to have a strong *inductive* reason for the claim that goodness is not analyzable. Given the failure of all attempted analyses thus far, goodness is *probably* unanalyzable in naturalistic terms. This, of course, is not a conclusive reason but it shows that Moore’s argument has force against analytic naturalism.

There are ways for naturalists to evade the Open Question Argument. Take the assertion that “X is good iff X is N.” As Moore (1903, 5) himself accepts, this assertion could be interpreted as (1) a claim about what things are good or as (2) a claim about what goodness is. Only the latter interpretation is relevant to the Open Question Argument since the argument can only rule out non-normative answers to the question, “What is goodness?” More specifically, the argument claims that ‘good’ and, say, ‘pleasurable’ cannot refer to the same property, not that they cannot refer to distinct and coextensive properties. Goodness may well be multiply realizable; that is, many different

²⁰ Cf. Finlay 2014, chapter 1.

things that are neither necessary nor sufficient for being good may characteristically cause or result from goodness. Similarly, many different things may characteristically cause or result from pleasure. Experiencing something beautiful, surprising, or humorous, or witnessing the achievements of a loved one are all things that characteristically cause pleasure, even though none of them is necessary or sufficient for being pleasurable. So, it seems that moral terms *can* be analyzed *synthetically* (rather than analytically) if we view moral properties as complex natural properties with robust causal profiles.²¹ This is what synthetic naturalists (or Cornell realists) claim.

Analytic naturalists base their claim that moral and natural terms refer to the same property on the alleged fact that they are synonymous. However, it has been argued that the assumption that supports the analytic naturalist thesis is wrong. The idea is that the synonymy between moral and natural terms is not necessary for them to pick out the same property.²² Take the assertion that “X is water iff X is H₂O.” Although ‘water’ and ‘H₂O’ refer to the same property, they are not equivalent in meaning. A competent user of the term ‘water’ need not know the fact that water is composed of molecules each of which include one oxygen atom bonded to two hydrogen atoms. Indeed, people did not know this fact before it was discovered in the 18th century. And it is not likely that the meaning of ‘water’ changed upon the discovery of the composition of water. Therefore, ‘water’ does not mean ‘H₂O.’ Rather, the meaning of the word ‘water’ is *causally regulated* by the tasteless, odorless, transparent liquid with relatively low viscosity that

²¹ Boyd 1988, 196–9.

²² Boyd 1988, 199–210; Brink 1989, chapter 6.

fills oceans, rivers, and lakes. Thus, saying “I know that X is H₂O, but is it water?” is *not* meaningless.

Likewise, saying “I know that X is pleasurable, but it is good?” is not meaningless because ‘goodness’ is not synonymous with ‘pleasurable,’ even though both refer to the same complex property. When we use the term ‘good,’ we refer to a functionally complex, not directly observable natural property that causally regulates our use of the term. There are many things that characteristically cause or result from goodness. We figure out which things are good by becoming aware of these characteristic causes and effects. It is possible for us to discover empirically that ‘being pleasurable’ is common to all good things. The *meaning* of ‘goodness,’ however, does not change upon this discovery because *it* is causally regulated by the complex property picked out by the terms ‘goodness’ and ‘pleasurable’. According to synthetic naturalists, then, “Good is pleasurable (or some other N)” is an *a posteriori* necessary truth like “Water is H₂O.” This causal regulation semantics help synthetic naturalists evade the Open Question Argument.

Causal regulation semantics gives synthetic naturalism *three* important advantages over analytic naturalism. First, synthetic naturalism successfully evades the Open Question Argument unlike analytic naturalism. Since statements such as “Water is H₂O” and “Good is pleasurable” are *a posteriori* necessary truths on this view, a synonymity between moral and natural terms is not required for them to pick out the same property. Thus, the Open Question Argument, at least in its most basic form, cannot refute metaphysical naturalism. It rather gives a reason to give up analytic naturalism. Second, synthetic naturalism captures the richness of moral experience better than analytic

naturalism. That goodness is multiply realized is an intuitive idea because actions and persons can be good in diverse ways. For instance, an action can be good because it is pleasurable, or beneficial, or enjoyable, and so on. There are also countless ways in which an action can be pleasurable, beneficial, or enjoyable. Non-reductive naturalistic metaphysics does justice to such variety. Third, synthetic naturalism promises a fully naturalistic epistemology unlike analytic naturalism. Since moral and natural terms are not equivalent in meaning, no conceptual analysis is needed to reveal the synonymy between them. Thus, *all* of the content of morality is empirically discoverable, according to the synthetic naturalist. Surely, such an epistemology fits better with the naturalistic aim of explaining moral practices by using empirical methods.

1.4 Problems

I have described various types of non-naturalism and naturalism. Each type of realism is subject to different objections. In this section, I discuss the problems associated with different forms of moral realism and determine a roadmap for the following chapters. My aim in this dissertation is to explore whether there is an alternative to moral realism that can capture what is most attractive about moral realism while addressing the objections that are brought about by that attractive feature. I have already argued in 1.2 that the existence of objective moral truths is the most attractive feature of moral realism. Thus, a plausible alternative to moral realism should retain morality's objectivity. In what follows, I will argue that such a view should also address two big problems of non-naturalism. First, it should account for the supervenience of the moral on the natural without having to offer a bloated ontology or collapsing into naturalism. Second, it

should address the epistemic challenge evolutionary debunking arguments pose to non-naturalism. Our alternative view should only deal with the objections to non-naturalism because, as I will show, it does not seem to be possible to retain the objectivity or genuine normative force of morality within a naturalistic framework. I will discuss these two big metaphysical and epistemic problems in the next chapters.

1.4.1 Moral Twin Earth

I ended the last subsection (1.3.3) by mentioning the advantages of synthetic naturalism over analytic naturalism. Synthetic naturalism has these advantages by virtue of its causal regulation semantics. However, the synthetic naturalist's semantics seems to be problematic. The shortcoming in causal regulation semantics is revealed by Horgan and Timmons's (1991) Moral Twin Earth Objection. First, imagine a world, a Twin Earth, where the chemical formula of a molecule of the tasteless, odorless, transparent liquid with relatively low viscosity that fills oceans, rivers, and lakes is ABC instead of H₂O. According to the causal regulation semantics, when the inhabitants of the actual world and the Twin Earth use the word 'water' they mean something different because the meaning of the word 'water' is causally regulated by different substances (ABC and H₂O) that play the same functional role.²³ Now, imagine a Moral Twin Earth, where the meaning of moral terms is causally regulated by different properties from those in our world. If causal regulation semantics is true, then moral terms such as 'right' or 'wrong' mean something else in Moral Twin Earth than they do in the actual world. However, this is counterintuitive. People in Moral Twin Earth might regard different kinds of actions as

²³ Cf. Putnam 1975.

right or wrong, but we don't think that they mean something different by 'right' and 'wrong.' Rather, we take this to be a substantive moral disagreement. But such disagreement is possible only if the inhabitants of the actual world and the Twin Earth mean the same thing when they use moral terms. Thus, there must be something wrong with the causal regulation semantics.

The Moral Twin Earth Objection raises two problems for synthetic naturalism. First, causal regulation semantics implies relativism. If causal regulation semantics is true, then it seems possible for different properties to causally regulate people's use of moral terminology in different societies or cultures. And empirically speaking, this seems to be true: it seems, for example, that different cluster properties causally regulate the use of the word 'good' in, say, Islamic theocracies and in modern liberal societies. But when they regard different kinds of actions as good, we think this shows a substantive moral disagreement rather than a difference in meaning of 'good.' If the meaning of moral terms differs in this way, then moral truth is relative to which cluster property causally regulates people's use of moral terminology in a certain society.²⁴ Second, the objection points to an epistemic problem for synthetic naturalists, who claim that moral facts just are natural facts. The question is, how do naturalists know which natural facts are the moral ones? Conceptual analysis and causal regulation semantics seem to be the only options available. Synthetic naturalists cannot choose the former option because they reject that moral and natural terms are synonymous. And if causal regulation semantics is problematic as Horgan and Timmons argue, then neither of these options is available for

²⁴ Of course, the fact that causal regulation semantics implies moral relativism will not bother those who have relativistic tendencies. Moral relativists such as Wong (2006) and Prinz (2007) are happy to endorse synthetic naturalism despite its relativistic implications.

synthetic naturalists, which means they have no way of establishing which natural facts are the moral ones.²⁵

Taken together, these problems seem to undermine the synthetic naturalist's project of explaining moral facts in terms of natural ones. The synthetic naturalist might attempt to provide a different account of moral language to avoid the implications of causal regulation semantics. This, however, would be quite difficult, considering the vital role causal regulation semantics plays in avoiding the Open Question Argument and in displaying the synthetic naturalists' methodological and epistemic commitments. As seen above, causal regulation semantics makes a fully naturalistic epistemology possible: if goodness is the property that causally regulates our use of moral terms, then we just need to empirically investigate what regulates the use of moral terminology to determine what goodness really is. But without such a semantics, it is not clear which natural facts are also moral. It seems, therefore, that offering a new account of moral language amounts to offering a completely new theory. And it is not clear whether this new theory can avoid the Open Question Argument.

There are criticisms of the Moral Twin Earth Objection. For example, Dowell (2016) argues that the objection is toothless because competent speakers' judgments about possible disagreement with hypothetical speech communities have no probative value for the development of a semantics for our moral terms. Such criticisms, however, pertain only to semantics. Thus, they are toothless against objections to metaphysical naturalism as such. The only difference between analytic and synthetic naturalism is that the latter argues that moral facts cannot be restated in non-moral or natural terms. But this is

²⁵ This is why the Moral Twin Earth Objection can be seen as an extension of the Open Question Argument. Cf. Bedke 2012.

merely a semantic point. According to the synthetic naturalist, some concepts and claims are irreducibly moral because moral *terms* are not synonymous with natural *terms*. Those terms still pick out the very same physical property. Metaphysically speaking, neither form of naturalism argues for ontologically unique, irreducible normative entities; that is, all moral properties and facts are also natural in the final analysis.²⁶ Hence, both versions of naturalism are subject to stronger forms of Open Question Argument that attempt to reveal the inadequacy of their moral ontology.

1.4.2 Normativity

I have argued that Moore's Open Question Argument cannot refute metaphysical naturalism. Nevertheless, there are more compelling forms of Open Question Argument that reveal the inability of metaphysical naturalism to capture the most distinctive feature of moral judgments, namely their *categorical* normative force. One influential version of this argument is called 'the Normativity Objection.'²⁷

Let's think about the relation between the natural features and the moral status of the following action: Mark steals his roommate Julia's money. Why is this action wrong? Is it because Julia needs that money to pay her rent? No, because the action would still have been wrong even if she did not need the money. Is it because Julia will be upset when she finds out? No, because the action would still have been wrong even if she would be happy to find out that her money got stolen. One could claim that the stolen money is her personal property. But now we just need to restate the question: why is theft of personal

²⁶ Cf. Sturgeon 1985, 239–40; Boyd 1988, 199.

²⁷ Parfit 2011; Scanlon 2014; FitzPatrick 2014.

property wrong? Is it because it is against the law? Again, why is it wrong to violate the law? A possible answer is that if a large enough number of people violate certain laws, this will make others feel less secure, and they will also break the laws to survive. In a Hobbesian state of nature, where there are no laws, no one will feel secure and thus people will do anything they think necessary for preserving their own lives. There will be no music, no paintings, no architecture. There will only be a constant war among people. One could also take an evolutionary perspective by claiming that violating laws will ultimately destroy the survival and reproductive success of individuals or humankind.

None of these proposals regarding the wrongness of Mark's stealing Julia's money is satisfactory because none of them can capture what we mean when we say Mark's action is wrong. We think that stealing would still have been wrong even if it did not give rise to a Hobbesian state of nature or reproductive failure. Why? Because we think that the moral wrongness of an action is independent of our desires, interests, or aims. That is, our moral judgments carry a categorical (desire-independent) normative force. The above proposals are *descriptive*, which can only capture non-categorical forms of normativity. Such descriptive statements can explain *instrumental* or *aim-relative* normativity, for example. If you want to survive, you should obey the laws. Descriptive statements can also explain *kind-relative* or *rule-relative* normativity. You can explain why a certain book is a *bad* book by stating that it has many missing pages. Or you can explain someone's being a *bad* speaker by pointing to the fact that they frequently misspell words. However, such descriptive statements fall short of explaining predicates such as being *morally* bad or having a *moral* reason to avoid something. The Normativity Objection states that descriptive or natural facts are "just too different" from moral facts

since they are facts about the physical nature of the universe and the causal laws that regulate the interactions of matter.²⁸ They can thus only explain normativity in the aim-relative, kind-relative, or rule-relative senses. But moral facts are normative in the *categorical* sense. In fact, categorical normativity is a non-negotiable element of moral discourse.²⁹ Thus, no naturalistic account of morality will be successful.

In fact, even in the case of artifacts, merely attributing some natural property to an object does not seem to fully explain *why* that object is good. For instance, we may talk about a good home alarm system. We can say that it is good because it has cameras, it has alarm monitoring capabilities, it has carbon monoxide detectors, and so on. But having these features does not make every artifact good; they may make another artifact bad. What is missing here is a reference to standards. This has been suggested by FitzPatrick (2008). According to him, evaluation has an implicit standard-based structure. When we say that a thing, a person, or an action is good or bad, we do not merely attribute natural features to them. Evaluation also involves a claim about whether having certain natural features *make* the thing in question satisfy the standards of goodness for that thing, i.e., a claim about whether such features are also *good-making* features. So, when we judge a home alarm system to be good, we not only describe its natural features, but we also think that the features involved in our description meet the standards of goodness for home alarm systems.

Similarly, when we say Mark's action is wrong, we not only describe the action but we also think that the complex set of natural properties involved in our description fail to meet the standards of goodness for human action. The important question is, of course,

²⁸ Enoch 2011. See also Paakkunainen 2018 for different forms of “just too different” objection.

²⁹ Cf. Joyce 2001, chapter 1.

whether the standards of goodness for human action are natural or not. In the case of artifacts, we can give a purely naturalistic characterization of standards. But given the categorical normative force of moral judgments, it is far from obvious that such a characterization is possible. The naturalist's real task is not to give some semantic story that enables them to avoid the basic version of the Open Question Argument. Rather, the naturalist must show that there are naturalistically characterizable action-guiding standards for human conduct *that are also categorically authoritative*. This could be an impossible task.

There are two main strategies to deal with the problem. The first strategy is to deny that moral facts are normative in the categorical sense. This amounts to rejecting our common conception of morality because those who adopt such a strategy define moral value in terms of what is *non-morally* good for the agent.³⁰ More specifically, they identify moral goodness with desires or responses of fully informed and coherent agents. In their view, the normativity of agents' non-moral good must be reduced to natural facts about what agents would desire themselves to desire if they had "complete and vivid knowledge of [themselves] and [their] environment." (Railton 1986, 174) Their idea is that we will *desire* the same thing or *respond* similarly to the various situations we might find ourselves in if the conditions of rationality are met: "the existence of reasons presupposes that under conditions of full rationality we would all have the same desires about what we are to do in the various circumstances we might face."³¹ I take this to

³⁰ Railton 1986, 173–5; Smith 1994, 202. There are, of course, other ways of naturalizing morality and normativity (or giving a reforming definition of them). For example, according to Copp's (2007) society-based account of morality, moral reasons are normative in the aim-relative sense. That is, moral goodness is what best serves the basic needs and interests of a given society.

³¹ Smith 1994, 198.

mean that, if we were omniscient and fully rational, we would share the general desire to do what is necessary to achieve one's aims in the most efficient way possible.³² Given this background desire and full knowledge, we would all act in the same way in relevantly similar circumstances.

One may want to call such a view a form of realism by claiming that the facts about one's constitution and circumstances are objective facts. What is important, however, is to determine where those facts get their *reason-giving force* from, because obviously they are not intrinsically reason-giving. The answer is, of course, desires. But desires are contingent; they change not only from person to person but also from time to time within individuals. What is more, the facts about our constitution and circumstances are contingent too, and they are mostly out of our control. So, even if we shared a background desire to be instrumentally coherent – or for that matter, even if we all had the exactly the same set of desires from birth to death – the facts about our constitution and circumstances would not provide the same reasons for everyone. Worse yet, considering the reason-giving force of the facts about one's constitution and

³² The thought cannot be that, for example, Mark would not want to steal Julia's money if he were coherent and fully informed about himself and his circumstances. Think about the following circumstances: some people abduct Mark's family, and they threaten him that they will kill his mom, dad, and his little sister if he fails to bring a certain amount of money within the specified time frame. They also tell him they will kill his family if he calls the police. Mark wants to save his family, but he does not have the money to do that. What should he do? Let's see what Mark and his ideal counterpart (Mark-Plus) could decide to do: (1) Mark suspects that the abductors will kill his family even if he brings the money. He believes that calling the police is the most reasonable thing to do in those circumstances. He calls the police, and the abductors kill his family upon hearing the police sirens. (2) Mark-Plus would know that calling the police would result in Mark's family's death. He also would know that stealing Julia's money would be the most efficient way for Mark to save his family, given the circumstances, and that Mark would successfully pull it off if he tried. Thus, Mark-Plus wouldn't want Mark to want to call the police. Instead, he would want Mark to want to steal Julia's money and then bring the money to the abductors to save his family.

circumstances, people could in principle have reasons to steal, kill, rape, torture, abduct, and so on, in certain extreme circumstances.³³ I believe Railton and Smith must bite the bullet and accept these highly counterintuitive conclusions.

The second strategy has been endorsed by synthetic and Neo-Aristotelian naturalists, and it consists of two steps: (1) subsume all normative concepts under one basic normative concept, and (2) argue that that basic normative concept refers to a natural property. For example, Schroeder (2005), following Scanlon (1998), takes the concept of a *reason* to be the basic normative concept. He then employs the Humean Theory of Reasons to show that the concept of a reason refers to a natural property. The Humean Theory of Reasons identifies reasons with desire-satisfaction. That is, Mark has a reason to steal Julia's money just in case stealing Julia's money will satisfy one of Mark's desires. This, if true, shows that reasons pick out natural properties. There is also a Neo-Aristotelian way of adopting this strategy. Foot (2001), for instance, takes the concept of *goodness* to be the basic normative concept. According to Aristotelian virtue ethics, goodness consists in what is conducive to a person's *flourishing*. And flourishing is understood as a complex natural property consisting of happiness, pleasurable, beneficence, healthiness, and so on.³⁴

The trouble with Railton's and Smith's "reforming" definitions of morality and normativity is that they eliminate the mystery of categorical normativity almost entirely simply by disregarding it. They say moral facts exist, but what they call 'moral' is not what we mean when we talk or think about morality. They just change the definition of

³³ The facts about one's constitution could generate reasons too. A sadist, for instance, could have a reason to torture people due to the facts about his constitution.

³⁴ This is similar to synthetic naturalism's definition of moral goodness.

morality so that they can provide a theory about how moral reasons exist. This is similar to changing the definition of God, saying that God is *not* omnipotent, omniscient, or omnibenevolent, and then proving its existence. Likewise, those who endorse the two-step reductionist strategy think that moral facts exist; however, they also think that moral goodness is relative to the facts about our desires or well-being. The second step is problematic because such naturalistic characterizations of the standards for good action can only explain normativity in one of the non-categorical senses outlined above. They thereby dismiss what is indispensable to morality, namely categorical normativity. Therefore, metaphysical naturalism does not appear to be well suited to explain the true nature of morality. *If* there are moral facts, they are to be explained by some form of non-naturalism.

1.4.3 Supervenience

While non-naturalism is better suited than naturalism to account for the categorical normative force of morality, non-naturalism also seems to be unable to explain the supervenience of the moral on the natural. In fact, I think this is a bigger problem for the non-naturalist than Mackie's queerness argument, although one can claim that the problem of supervenience is an extension of the queerness argument.

According to Mackie's (1977, 38–42) famous argument from queerness, moral properties, if they exist, have two important features: (1) they inescapably motivate us to comply with their demands, and (2) they provide us with universal and attitude-independent (or categorical) reasons for compliance. Mackie thinks that such properties or values must be utterly different from anything else in the universe, i.e., they must be

queer. The idea is that since moral properties or objective values cannot be detected by the five senses or by any scientific instruments, there are no moral properties or objective values.³⁵

I don't think this is a very strong argument since it basically says something is queer and thus it does not exist. If queerness amounts to being different from the things that natural sciences study or not being detectable by empirical methods, and if the argument from queerness is correct, then things such as space, time, fields, numbers, or consciousness cannot exist because they are queer in the same sense. The queerness argument proves too much because it implies that no entity that seems different from other things can exist.

The assumption behind Mackie's argument must be that we should reject the existence of any entity that fits poorly with our best scientific account of reality. This is a controversial naturalistic assumption that is rejected by non-naturalists. For example, should we reject mathematical and logical truths altogether on the grounds that mathematical and logical properties cannot be described empirically?³⁶ This sounds more 'queer' than the non-naturalist accounts Mackie is attacking. Supervenience, however, is a thesis that both naturalists and non-naturalists agree on because it is not based on our scientific worldview but rather on common sense. Therefore, it is regarded as one of the

³⁵ Mackie's reasoning leads to an error theory. Error theorists believe that since there are no objective moral values, and since our moral judgments purport to describe objective moral facts, all of our moral judgments are false.

³⁶ Mackie is optimistic about giving satisfactory empirical accounts of things such as essence, number, identity, diversity, substance, necessity, possibility, causation, and so on. However, he says that if it turns out that these things cannot be explained in empirical terms, "then they too should be included, along with objective values, among the targets of the argument from queerness" (Mackie 1977, 39).

least controversial theses in metaethics. The task of explaining the necessary dependence of the moral on the natural is a more pressing problem for non-naturalists than the queerness argument.

A possible solution to supervenience is to postulate a separate Platonic universe of values which somehow affects our universe. It is, however, hard to accept such a view. First, it is not ontologically parsimonious. It is not clear at all whether we really need a Platonic normative heaven to account for moral objectivity. Second, how are we to explain the interaction between two distinct universes? We have no idea how they are supposed to interact or why they should interact in the first place unless, of course, we want to appeal to a divine plan or a pre-established harmony. We probably will need additional laws of nature which explain the relation between our universe and the Platonic normative universe. Can we ever know whether such laws exist? Can we reveal them even if they exist? The Platonic option might solve the problem of supervenience, but it creates more problems than it solves.

Shafer-Landau (2003, 76–8) attempts to explain supervenience in non-reductionist terms. As I mentioned in 1.3.1, he espouses the exhaustive constitution thesis, according to which moral properties are exhaustively constituted by natural properties, even though moral properties cannot be reduced to or analyzed into their more basic constituents. That is, the non-identity of a given moral property with any single natural property or a set of natural properties is consistent with the constitution of every moral property by some series of natural properties. So, Shafer-Landau seems to think that moral properties are emergent properties.³⁷

³⁷ Emergence occurs when some features of the world are not reducible to any of their parts nor to any of the relations between their parts. Emergent properties arise out of more fundamental

However, the exhaustive constitution thesis alone does not entail supervenience. A further premise is needed to explain supervenience. Shafer-Landau further claims that there could be a *metaphysical entailment* between moral and natural properties. Even though it is not a conceptual matter that some natural property N grounds some moral property M, it could be a conceptual truth that *if* N grounds M in one instance, then N necessarily grounds M in other instances. For example, it could be a conceptual truth that *if* table salt is NaCl then it always is. If exhaustive constitution and metaphysical entailment theses are true, then the supervenience of the moral on the natural should not surprise us. Two naturalistically identical actions (intra-world supervenience) or universes (global supervenience) must be also morally identical because moral properties are (i) fully constituted and (ii) metaphysically entailed by natural properties.

In Shafer-Landau's (2003, 74–5) view, then, the moral status of an action has its source in the natural features of the action. Think again about Mark's action. Let's say he steals Julia's money because she often is passive aggressive towards him, and he wants to punish her because of that. Also, his favorite rock band will perform at a festival in his city in a week, but tickets are quite expensive. Since buying a ticket out of pocket will certainly put a strain on his finances, he thinks stealing Julia's money is an attractive option. Now, what is the relation between these descriptive facts related to Mark's theft and the fact that his action is wrong? It seems that we cannot explain the nature of his committing theft at all if we take away all of these descriptive facts. So, natural facts at

properties and yet are irreducible to these fundamental properties. Quantum entanglement and covalent bonding are among the examples of emergence in science. Quantum entanglement has the special feature that when two objects combine, the state of the joint object is not determined by the states of the original objects. Covalent bonding has the special feature that when two atoms bond, electrons lose their individual identities and stop being the components of the joint object, which has one joint charge.

least partly constitute moral facts. Shafer-Landau makes the stronger claim that moral facts are *exhaustively constituted* by natural facts³⁸ because if they are not and if we reject reductionism, then the only remaining metaphysical option is substance dualism³⁹ and the only remaining explanation of supervenience is divine intervention.⁴⁰

There are objections to Shafer-Landau's metaphysical entailment thesis, which I cannot review here.⁴¹ Even if we grant the metaphysical entailment thesis, there are problems with the exhaustive constitution thesis. First, as FitzPatrick (2008, 190–4) argues, the exhaustive constitution thesis misses the role of standards in determining moral facts. Even if we describe all the facts related to Mark's theft, our description does not address why these natural facts *make* his action wrong. (Recall the home alarm system example.) Merely listing the natural features of Mark's action does not tell us why those features are *wrong-making*, i.e., why they fail to meet the standards of goodness for human action. Second, the exhaustive constitution thesis commits Shafer-Landau to metaphysical naturalism of the kind supported by Cornell realists. This metaphysical commitment may enable him to explain supervenience, but it also makes him share the same burden of explaining normativity as Cornell realists (cf. 1.4.2). It is not clear how

³⁸ "There is nothing to a case of generosity, or viciousness, or dutiful action, other than the natural features that constitute such properties. Something exemplifies a moral virtue entirely in virtue of its possessing certain natural features" (Shafer-Landau 2003, 75).

³⁹ Substance dualism is a view in philosophy of mind, according to which reality consists of two distinct and independent ontological realms: the physical and the mental. In moral ontology, 'physical' and 'mental' correspond to 'natural' and 'moral,' respectively.

⁴⁰ "If the moral is neither identical in kind to the natural, nor exhaustively constituted in particular cases by it, then there is no explanation (other than God's fiat) of why the moral should be specially dependent on the natural as it is" (Shafer-Landau 2003, 78).

⁴¹ See Mabrito 2005 and Ridge 2007 for objections to Shafer-Landau's treatment of supervenience.

metaphysical naturalism is compatible with Shafer-Landau's commitment to intrinsic (or categorical) normativity.⁴²

There is an ambiguity in Shafer-Landau's statements about non-reductive naturalism. On the one hand, he accepts that he adopts the same naturalistic moral ontology as Cornell realists. He says the only difference between his non-naturalism and non-reductive naturalism is methodological and epistemological.⁴³ On the other hand, he thinks non-reductive naturalism corresponds to 'property dualism' in philosophy of mind.⁴⁴ According to property dualism, there are two distinct ontological kinds of properties, namely physical and mental properties; however, mental properties arise out of physical ones. Shafer-Landau's latter statement seems to misidentify non-reductive naturalism because Cornell realists (or synthetic naturalists) do not argue for two ontologically distinct kinds of properties. Their non-reductionism is solely a semantic one (cf. 1.4.1). That is, while reductionist accounts render moral explanations dispensable, such explanations acquire a special status under non-reductionist accounts. Just as biologists or geologists cannot and do not explain biological or geological concepts in terms of the interactions between elementary particles in physics, so philosophers cannot and do not use the language of physics when they talk about moral concepts. Be that as it may, there is only *one* ontological kind of properties: all moral properties and facts are

⁴² Cf. Shafer-Landau 2003, 206–7.

⁴³ Shafer-Landau 2003, 63–4.

⁴⁴ Ibid.

also *natural*. To put it differently, we will only see natural properties on an ontological X-ray screen.⁴⁵

Let's put this misidentification problem aside, and grant, for the sake of the argument that synthetic naturalism is a form of property dualism. What would be the implications of such a view? Since property dualism talks about two distinct ontological kinds of properties, the universe has an irreducibly and uniquely normative dimension, on such a view, even though all normative properties and facts arise out of natural properties and facts. This poses two problems. First, property dualism allows for two types of instantiations. While natural properties are instantiated by entities of the same kind, moral properties are instantiated by entities that are of a different ontological kind. What secures the connection between properties of different ontological kinds? The idea of a "cross-ontological-realm" instantiation sounds mysterious.⁴⁶ Second, what makes the interaction between properties of different ontological kinds possible? Property dualism gives natural properties an ontological priority because moral properties are dependent on (yet irreducible to) natural ones. Combining a 'priority claim' with a 'non-reductionist claim' makes it difficult to understand how the moral, which is a mere byproduct of the natural, can acquire a distinct ontological status and still *affect* the natural. I believe we might need to appeal to a supernatural entity, such as God, to explain such interactions. Property dualism seems to cohere better with Shafer-Landau's commitment to intrinsic

⁴⁵ For example, Sturgeon, a Cornell realist, writes, "[a]s a philosophical naturalist, I take natural facts to be the only facts there are" (1985, 239). Boyd, also a Cornell realist, writes, "[t]he moral realist may choose to agree that goodness is probably a physical property but deny that it has any analytic definition whatsoever" (1988, 199).

⁴⁶ Cf. Benovsky 2015, 4.

normativity,⁴⁷ but it might come at the cost of losing the advantage of explaining supervenience.

Shafer-Landau also seems to overlook a possible non-naturalistic explanation of supervenience. He says that we must explain supervenience by appealing to God's commands if we reject the exhaustive constitution thesis and reductionism. This must be because he thinks that the only remaining metaphysical option would be substance dualism, and only a divine intervention could account for the mysterious interaction between two ontologically distinct substances. However, one could also endorse 'dual-aspect monism.' According to dual-aspect monism in philosophy of mind, there is only one type of substance with two different aspects: physical and mental. That is, the person has both physical and mental aspects, which are irreducible to each other. These aspects are accessible to us in different ways. For example, the headache I am experiencing right now is accessible to me in a mental (or qualitative) way. If I were hooked up to a machine that could monitor my brain activities in the finest detail, my headache would be accessible to me in a physical way. There is no ontological priority on this view. The fundamental substance is neither physical nor mental. Rather, we can call it 'physical-mental' or perhaps *phental*.⁴⁸

⁴⁷ This is only because we have seen that naturalistic accounts are unable to account for categorical normativity. Yet it is doubtful, as I will argue later, that we must posit irreducibly normative entities to show that morality is objective or categorically normative.

⁴⁸ I borrow this term from Benovsky (2015). The duality here is like wave-particle duality in quantum physics. Experiments such as double-slit experiment show that elementary particles behave like particles under certain experimental conditions, whereas they behave like waves under others. The particle-like nature and the wave-like nature of a quantum object are two different *aspects* of the same entity, which we can call a 'wave-particle' entity or a *wavicle*.

FitzPatrick (2008, 2018) takes up a dual-aspect view of moral ontology and explains supervenience within that framework. First, he agrees with Shafer-Landau that natural features of an action fix its moral status. Artifacts or actions are good or bad in virtue of their natural features. We cannot evaluate anything if we take away their natural features. Let's say Julia is playing the Godfather theme song on her guitar while planning revenge on Mark for stealing her money, and let's say we want to make a judgment on the quality of what she is playing. Certainly, if we take the guitar out of the picture, it becomes impossible to make an evaluation because there will be nothing to evaluate. The same goes for Julia: without her, the thing to be evaluated, namely 'the Godfather theme song played by Julia on guitar,' will disappear. Songs or melodies do not float free, bearing no relation to composers, musicians, singers, musical instruments, instrument-makers, and so on. Similarly, moral properties do not float free, bearing no relation to natural properties.

Second, as discussed above, he says that merely listing the natural features of a thing does not count as evaluation. We must also explain *why* these natural features *make* the thing good or bad. To do that, we must determine whether or not they meet the standards of goodness for that thing. For example, if we take the standard of goodness for playing the Godfather theme song to be a particular sequence of musical notes with different pitches and durations, our evaluation of Julia's performance will depend on how well the sequence of the notes played by her match *that* specific sequence. Likewise, moral evaluation consists both of describing natural features of an action and determining whether those features meet the standard of goodness for human action.

In FitzPatrick's view, Shafer-Landau's exhaustive constitution thesis is superfluous because the necessary dependence of the moral status of an action on the action's natural features, in the sense described above, is sufficient for explaining supervenience.⁴⁹ The crucial claim here is that the natural features of an action fix its moral status regardless of the metaphysical status of the moral standards.⁵⁰ According to FitzPatrick's dual-aspect view, moral standards are derived from irreducibly evaluative aspects of the reality that we call 'natural.' That is, an action – like everything else in the universe – has both natural and evaluative aspects that are irreducible to each other. On the one hand, there are properties and facts that we typically call 'natural.' We discover them through empirical investigation. On the other hand, these very same properties and facts are also *inherently value-laden*, just as particles are also waves. The evaluative dimension of properties and facts cannot be detected by scientific instruments; rather, they are to be grasped only by engaged and properly informed moral experience and reflection.⁵¹ Although it is these evaluative aspects of reality that generate the correct moral standards, actions or persons are still good or bad *by virtue of* their natural properties. In other words, even though we cannot *explain* moral standards simply by appealing to empirical (and familiar) *aspects* of human life, moral standards are still rooted in these familiar

⁴⁹ FitzPatrick 2008, 190–8.

⁵⁰ “[A]s long as individuals meet or fail to meet the standards that apply to them simply by virtue of their natural properties, it makes no difference what the metaphysical status of those standards is; once the standards are in place, supervenience will be preserved, just as it is for artifacts” (FitzPatrick 2008, 197).

⁵¹ *Ibid.*, 195.

properties and facts, in the sense that they are determined by the *irreducibly evaluative aspects* of the world.⁵²

FitzPatrick thinks that the supervenience relation between the moral and the natural holds independently of the metaphysical characterization of the moral standards. To see this point, imagine that ‘the musical standard’ goes over and above merely imitating songs, such that “If S instantiates the non-musical property FW (e.g., frequencies and wavelengths), it also instantiates the musical property M.” That is, there are irreducibly musical properties that *make* certain sequences of notes *musically good*. Although musical goodness would be stance-independent and *non-natural*, Julia’s performance would meet or fail to meet the musical standard *by virtue of* the *natural* properties and facts related to her performance. Nothing can instantiate M if we take away the natural properties and facts leading to the production of FW. Two naturalistically identical songs or universes will be also musically identical. Thus, supervenience holds.⁵³

The first thing to notice is that we must assume a fundamental metaphysical relation between moral and natural properties for FitzPatrick’s account to work. That is, this way of explaining supervenience requires not only to espouse Shafer-Landau’s metaphysical entailment claim but also to assume a fundamental association of particular evaluative features with particular natural features. FitzPatrick’s background assumption is not only that an adequate account of objective morality must be non-naturalistic, but also that we

⁵² An ‘aspect’ is different from a property because it is not instantiated by actions or persons. ‘Aspect,’ in this context, is “a way to describe” an entity (Benovsky 2015, 9). It refers to *how* an entity *is*, or how it appears to be.

⁵³ “[E]thical standards are derived from the set of values inherent in a subset of the properties and facts we also identify as natural ones, and this means that, once the natural facts are fixed, so are the ethical ones” (FitzPatrick 2008, 198).

must posit irreducibly evaluative or normative properties and facts to adequately explain the objective (or categorical) aspect of morality.⁵⁴ I have argued in 1.4.2 that moral facts, *if* they exist in the way we think, should be explained by some form of non-naturalism. However, this does not entail that certain features of the world are value-laden in the way FitzPatrick describes. Do we really need these extra, purely evaluative aspects to account for moral objectivity? FitzPatrick himself admits that the idea of a fundamental ontological connection between the moral and the natural creates a mystery that cannot be unraveled by philosophical investigation.⁵⁵ So, we should not assume such fundamental metaphysical associations unless FitzPatrick's dual-aspect view is the *only* plausible explanation available to account for objectivity or categoricity. In fact, even if turns out that the dual-aspect view provides the only appropriate moral ontology to explain objectivity, I don't think error theorists should be blamed if they reject outright such an inflated ontology and say that morality does not exist. In any case, I believe it is possible to offer a non-naturalist account of objectivity with no commitment to an inflated ontology of the sort FitzPatrick proposes, if we support our account with a plausible *constitutivist* origins story. The primary aim of the Chapter 3 will be to discuss such an account.

The second thing to notice about FitzPatrick's dual-aspect solution to supervenience is that it fails to address the main intuition behind the supervenience thesis. According to

⁵⁴ Cf. FitzPatrick 2018.

⁵⁵ "So if these sorts of connections are among those that people find mysterious when they speak of supervenience, then it is true that I have done nothing to lessen the mystery. My view is that certain elements of the world just are value laden in this way, as a basic metaphysical fact about them, and that there may not be anything more for philosophy to say here" (FitzPatrick 2008, 197).

the supervenience thesis, there is a *one-way* necessary dependence: moral properties necessarily depend on natural properties, *not vice versa*. That is, natural properties can change without there being a change in moral properties (cf. 1.3.2). Shafer-Landau's exhaustive constitution thesis (minus property dualism) is well suited to explain supervenience because it *prioritizes* natural properties over moral ones. However, natural properties do not enjoy an ontological priority on FitzPatrick's dual-aspect view. Fundamental entities of the universe are not purely natural, nor they are purely evaluative under dual-aspect monism. Rather, the natural and the evaluative are two different *aspects* of the *same* ontological kind of entity, which we can call 'evaluative-natural' or *evalural*. So, if "the very features we have been calling 'natural' all along [...] were never *merely* natural to begin with,"⁵⁶ then it seems that there can be no change in the natural features of the world without there being a change in the moral features of the world.

What does it mean to say, on a dual-aspect view, that natural features of an action fix its moral status? A change in the natural aspect of a thing *must entail* a change in the evaluative aspect of it, provided that there is nothing purely natural or evaluative in the world (i.e., everything is *evalural*). To deny this would amount to abandoning the dual-aspect solution altogether. After all, the assumption of this fundamental ontological connection between the natural and the evaluative is what enables FitzPatrick to explain the *covariance* between the natural and the moral. Surely, if we regard supervenience only as covariance, then such a solution might work. However, I believe what makes the supervenience thesis so commonsensical is the thought that natural properties constrain

⁵⁶ FitzPatrick 2008, 195.

moral properties, *but not vice versa*. Intuitively, a change in the natural features of an action does not necessarily entail a change in its moral status. It is always possible that some natural feature is morally irrelevant. For example, if it is wrong for Mark to steal Julia's money, Mark's action would still be wrong even if the entrance door to their house had a different color. It is hard to understand how the natural can have this ontological priority on the dual-aspect view.

How can it be that a descriptive change does not entail a moral change when the evaluative and the natural are literally the same thing? First, you may say that there are only *three* results of a moral evaluation, and that the result does not change unless there is a big enough change in the natural properties. In other words, an action is either (1) morally good/right, (2) morally bad/wrong, or (3) morally neutral/permissible, and minor descriptive changes may not be sufficient for a transition to occur between these states. The trouble with this objection is that it suggests a rigid picture of morality. There are countless good and bad actions, and in many cases, we can evaluate them against each other. It is not that all good or bad actions have the same moral status. Given the richness of moral experience and evaluation, it is likely that moral goodness comes in degrees and that there are countless moral states between the two extremes of the moral spectrum, just as there are infinite number of colors. So, every descriptive change must correspond to an evaluative change in the same thing under dual-aspect monism, no matter how difficult (or impossible) it is for us to discern such differences.

Second, you may say that a change in the evaluative status of a thing does not entail a change in the *moral* status of that thing. For example, a coffee maker that cannot make coffee is a bad coffee maker. It can become a good coffee maker if we fix it. However,

this of course would not also make the coffee maker *morally* good. Even though the evaluative is a more general category than the moral, this objection does not fully overcome the problem. If properties we have been calling ‘natural’ were never merely natural, but they are also evaluative, then a descriptive change must still correspond to an *evaluative* change. But this, if true, has highly counterintuitive consequences. For example, let’s say we are evaluating two beaches, namely beach A and beach B. A and B are naturalistically identical except that in A there are 64×10^{18} sand grains, whereas in B there are $64 \times 10^{18} - 1$ sand grains. This descriptive change does not entail an evaluative change: if A is a good beach, then B is also a good beach. It might thus not be a good idea to employ the dual-aspect strategy to characterize the normative domain.

The reason all this talk about the metaphysical relations between natural and moral properties sounds mysterious and somewhat odd could be the conviction that moral properties exist ‘out there’ (or they are tied to human rationality as an absolute inner value), i.e., they are part of the fabric of a stance-independent reality. It seems true that moral objectivity cannot be captured by naturalistic explanations. But it also seems true that modeling moral ontology on scientific ontology leaves us with an unsolvable mystery. On the one hand, we are familiar with empirical properties and facts; we know, more or less, what they are and how they are related to each other. On the other hand, we don’t know whether non-natural properties and facts exist. We don’t know how they are supposed to be and how they are supposed to relate to each other and to other empirical properties if they exist. We don’t know how they are supposed to make morality objective. We don’t even know whether we can ever know that they exist. Moral realists think they have no option but to postulate these non-natural entities because they think

there is no other way to account for morality's objectivity. However, the realist strategy not only poses seemingly unsolvable metaphysical problems but also runs the risk of conflating the kind of objectivity possessed by empirical facts with the kind of objectivity possessed by moral facts. Empirical and moral facts are fundamentally different. While empirical facts are about 'what *is the case*,' moral facts are about 'what *ought to be done*.' Given this fundamental difference, it would not be very surprising if it turned out that moral facts had a different metaphysical status than empirical facts, and that they possess different kinds of objectivity.

The difference could be that moral properties and facts, unlike empirical ones, do not exist 'out there' and wait to be discovered by us (or they are tied to human rationality as an absolute inner value). Instead, they could *exist* in a *non-ontological* way like mathematical and logical facts, as Parfit suggests (cf. 1.3.1). Mathematical or logical truths are not based on accurate descriptions of the world; rather, they are *a priori* truths that hold independently of our desires, interests, or aims. Moral objectivity could be explained in a similar fashion, should moral truths be *a priori* like mathematical and logical truths. Moreover, if moral properties and facts exist in a non-ontological way, then there is no need to tell some metaphysical story about the relation between two ontologically different kinds of entities that are nevertheless necessarily coextensive. The ontological priority of the natural will not pose a problem either, because the Parfitian view does not conceive of the world as consisting of ontologically distinct entities to begin with.

However, Parfit's non-ontological view is as mysterious as other non-naturalistic views if we do not complement it with a plausible *origins* story. The mere claim that

moral truths have the same non-ontological *a priori* status as mathematical and logical truths does not tell us much, because it leaves many questions unanswered. For example, why is morality ontologically similar to mathematics and logic, given that morality is fundamentally *practical* unlike these fields? What is the *source* of morality? (Or where do moral principles come from?) What makes moral principles objective? How is it possible for something purely formal to generate substantive moral content? In Chapter 3, I will explore these questions and offer a *Kantian constitutivist* origins account, which conceives of moral goodness as nothing more than the way our reason necessarily functions. On this *stance-dependent* account, the rules of mathematics, logic, and ethics are provided by our own reason.⁵⁷ It is not that we have to first learn the basic rules of logic, and only then we will be able to apply them later. Rather, we must already be applying the fundamental laws of logic to be able to learn logic at all. For example, even little children who never have attended a logic class can understand what a contradiction is. This is because the law of contradiction partly *constitutes* how we all think.⁵⁸ And the same goes for the basic rules of mathematics. Nevertheless, there is an important difference in the case of ethics. While mathematical and logical laws are constitutive of how we think theoretically, moral laws are constitutive of how we think *practically*. (This difference surely needs an explanation.) According to this version of Kantian constitutivism, objective moral facts with genuine normative force are grounded in our reason in the sense that moral laws are constitutive of how our practical faculties operate.

⁵⁷ Even though these rules exist independently of our desires, interests, aims, or conscious decisions, they do not exist independently of the *standpoint of reason* because such rules are tantamount to how our reason naturally functions.

⁵⁸ “The laws of logic govern our thoughts because if we don’t follow them we just aren’t thinking” (Korsgaard 2009, 32).

1.4.4 Moral Epistemology

Conceiving of moral properties as entities that exist ‘out there’ (or as tied to human rationality as an absolute inner value), as part of the fabric of the universe also raises serious epistemic issues. Some of these issues can be sidestepped by adopting naturalism, but as argued above, the naturalistic project fails to account for the most distinctive feature of morality, namely categorical normativity. Since my aim in this dissertation is to explore an alternative to moral realism that can capture the objectivity and normativity of morality, I will not resort to naturalism to solve these daunting epistemic problems. Rather, I will claim that the perceptual characterization of moral knowledge is not necessary for non-naturalism. If we stop regarding moral properties as non-natural entities that are waiting ‘out there’ to be discovered by us (or as non-natural entities that are tied to human rationality as an absolute inner value), we can circumvent certain important epistemic problems associated with non-naturalism. However, all forms of non-naturalism seem to be subject to evolutionary debunking arguments, which raise doubts about the existence of moral knowledge.

There are two important questions regarding the knowledge of stance-independent moral properties. First, how can we know that a particular property is moral? Second, is it possible to know anything about moral properties besides the fact that they are moral? It seems that the naturalist can easily answer the latter question because moral properties just are natural properties on their account, and thus we can know them through scientific investigation. The former question is harder for the naturalist to answer because even if we grant that moral properties are natural properties, that does not tell us *which* natural

properties are also natural. You might think that moral properties are identical to properties we are familiar with, such as the property of being honest or the property of being pleasurable. But the problem is that questions such as, “Is it good to be pleasurable?” or “Is it morally right to be honest?” are open questions (cf. 1.3.3); so, we cannot easily identify certain natural properties with certain moral properties. If we knew, for example, that moral goodness consisted in maximizing pleasure, we would be able to apply ordinary scientific methods to ascertain whether an action is morally good. But it seems that we cannot vindicate the truth of the naturalistic principle that links goodness with maximizing pleasure in an empirical way.

This identification problem can be overcome by embracing causal regulation semantics. If our use of moral terminology is causally regulated by complex natural properties picked out by terms such as ‘goodness’ or ‘rightness,’ then we can in principle know which natural properties are also moral. For example, we know that our use of the word ‘water’ is causally regulated by the tasteless, odorless, transparent liquid with relatively low viscosity that fills oceans, rivers, and lakes. Thus, we can say that a certain set of natural properties corresponds to the word ‘water.’ Similarly, if we want to figure out which natural properties are also moral, we need to figure out which properties causally regulate our use of moral terms. We need to ask ourselves what kinds of actions or character traits make us say things like, “He didn’t do the *morally right* thing,” or “She is a *morally good* person.” Of course, moral rightness or goodness is exemplified in various different ways under this view (cf. 1.3.3); however, we can still figure out what *kinds* of complex natural properties correspond to ‘rightness’ or ‘goodness.’

However, one might say that the identification problem is not only a semantic problem, but it is also an ontological one. The idea is that ‘goodness’ or ‘rightness’ is *intrinsically normative*; that is, moral terms carry *categorical* reason-giving force that cannot be captured by scientific descriptions of the world. Take the moral principle that links goodness with maximizing pleasure. We cannot determine the truth of this principle empirically because such an investigation will only tell us what *is* the case rather than what *ought to be* the case. Scientific investigations will reveal facts about the physical structure of the universe and the causal laws that govern the interactions of matter. But no matter how detailed our scientific descriptions are (or will become), we will not be able to scientifically confirm the existence of categorical normativity, for interactions between atoms and molecules do not tell us why we *should* behave in certain ways rather than others. Categorical normativity, however, is indispensable to morality (cf. 1.4.2). So, we must not rely on scientific methods if we are to have moral knowledge.

The problem is that once we leave the methods of natural sciences behind, it becomes unclear *how* we are supposed to acquire moral knowledge. If scientific methods are not an option, then the only way to attain moral knowledge seems to be through *intuition*. Intuitions are immediate intellectual grasps or appearances prior to reasoning. To have an intuition is to have a mental state in which something *seems true* to you before you advance an argument for it. In other words, you have an intuition about X when you *immediately* become aware of the truth of some aspect of X upon reflecting on it, i.e., something about X seems true to you before you start reasoning about it.⁵⁹ We have

⁵⁹ We can have various kinds of intuitions. While “2 is greater than 1” is a mathematical intuition, “the snow is either white or not white” is a logical intuition. Notice that our having these intuitions does not depend on reasoning or argumentation. We also have intuitions when we play

moral intuitions when something seems *morally* right, bad, permissible, and so on. More specifically, to have a moral intuition, according to many non-naturalists, is to perceive non-natural moral properties attached to actions, persons, or character traits. The idea is that we are justified in believing some of our moral statements when we adequately understand their content or when their content strikes us as true. Many non-naturalists are also moral intuitionists because they dismiss empirical methods as unreliable in attaining moral knowledge.⁶⁰

If moral intuitionism is true, then it seems easy to answer the first question mentioned above, namely, how can we know that a particular property is moral? If moral goodness or rightness is a non-natural property, then being immediately aware of it could give us sufficient evidence for claiming that it is moral. It is not easy, however, to determine what we can know about moral properties apart from the fact that they are moral. Admittedly, if it is true that we directly perceive moral properties through intuition, then we do not only directly perceive that they are moral, but we also directly perceive other possible features of them. But can we say anything about how such a mechanism is supposed to work? We have plausible accounts of how sensory perception connects us to the world. In fact, even before the development of modern scientific understanding of, for example, the visual system, people knew that we get images of the world through our eyes. They knew that our eyes are on the outside of our heads and provide the connection between us and the external world. The same goes for other sensory systems. By contrast,

games. For example, in bullet chess games where each player is given only a minute in total, players mostly play by their 'chess intuitions' instead of active thinking.

⁶⁰ Some prominent examples of moral intuitionism are G.E. Moore (1903), W. D. Ross (1930), Shafer-Landau (2003), Audi (2004), and Huemer (2005).

nobody seems to know how intuitions could give us access to facts. We don't really know what an intuition consists in. Does it involve some part of the brain? If yes, which parts of the brain are active when we have a moral intuition? More importantly, how do non-natural moral properties *cause* our intuitions? It is not clear whether we can ever know the answers to these questions. Intuitionist epistemology, like non-naturalistic accounts of supervenience, leaves us with an unsolvable mystery.

Perhaps 'mysterious' is not the right word here. After all, I have claimed in 1.4.3 that denying the existence of something just because it seems different from things we are familiar with is not a good strategy because it proves too much. In fact, I am sympathetic to Michael Huemer's (2001) phenomenal conservatism, according to which *appearances* are the source of justification for our beliefs.⁶¹ That is, it is rational to assume that things are the way they appear unless and until you have reasons to the contrary. For instance, if it seems to us that morality is objective and categorically normative, we don't need a reason to believe that morality has such a status; we can rationally keep presuming the truth of objectivity and normativity in morality until we have a reason to doubt it. (I will argue in the following that evolutionary debunking arguments give us reason to doubt

⁶¹ There are at least four types of appearances: (1) sensory perceptions, (2) memory experiences, (3) introspective appearances, and (4) intuitions (cf. Huemer 2001, 98–108). Huemer's idea is roughly that our beliefs are based on what seems true to us whenever we seek truth. This is true not only of reasoning or argumentation (e.g., premises or steps in an argument seem true to us), but also of epistemic beliefs (e.g., foundationalism seems true to us). To reject that appearances are sources of justification would be *self-defeating* because that belief itself is an appearance (ibid., 105). One must admit that certain appearances *do* confer justification to avoid self-defeat, but then they also must accept that those justification-conferring appearances are the ones that we don't have specific reasons to doubt. And this is exactly what the phenomenal conservatist argues (ibid., 106–7).

that morality is objective and categorically normative.)⁶² Similarly, we can rationally keep presuming that we grasp moral truths through intuition until we have grounds for doubting it. Merely saying that moral intuitions are mysterious does not give us enough reason to reject non-natural moral properties and facts, because things can exist even though we cannot describe them empirically.⁶³ However, appearances are *defeated* when they conflict with other appearances that are epistemically more trustworthy. In other words, we should not trust appearances if we have grounds for thinking that they are unreliable. Even though queerness or mysteriousness is not a sufficient reason to reject the existence of moral intuitions, there are various reasons to believe that moral intuitions are unreliable.

First, it is improbable that our intuitions would reliably track moral facts. Intuitions seem to be part of the natural world since they consist of psychological states. That is, whenever we have an intuition, we are in a certain mental state that is caused by some event or state of affairs in the world. But if moral properties and facts are *causally inert*, i.e., if they are causally isolated from our mental states, how is it that they bring about a cognitive state in which something seems true to us? It is not clear at all how intuitions could reliably track moral facts given that such facts are causally disconnected from our intuitions. Thus, Bedke (2009) says that it would be a “cosmic coincidence” if our intuitions reliably tracked moral facts. By contrast, it is not problematic to say that our sensory perceptions reliably track natural facts because the reliability of our sensory perceptions is secured by appropriate causal connections. For instance, the fact that I see

⁶² That said, I don’t believe that such debunking arguments give us strong enough reason to abandon the idea that morality is objective and categorically normative. For more on this, see 2.2.

⁶³ For example, consciousness, mathematical and logical truths, numbers, space, time, and so on.

a rabbit in my living room at the moment is a good justification for my belief that there is a rabbit in my living room, provided that there is nothing unusual about my visual system and that nothing affects its proper function. We have no grounds for thinking that our moral intuitions have such justificatory force since the perceived ‘moral object’ seems to lack any causal power.

Although we cannot say for sure that moral properties and facts lack causality, we have good reason to think so. Let’s think about what sort of causal powers moral properties might have. What would cause your negative reaction if you saw, for example, someone torturing a rabbit on the street? Let’s call the natural properties associated with that action ‘ N_A ,’ and your reaction ‘ N_R .’ There is also a moral property, namely wrongness, that supervenes on N_A . Let’s call that property ‘ W .’ We seem to have three options: (1) your reaction, N_R , is caused exclusively by N_A because W is causally impotent; (2) your reaction is caused exclusively by W ; (3) your reaction is caused both by N_A and W . If we choose the second option, then we reject outright the causal closure of the natural world⁶⁴ and commit to the belief that moral properties ‘float free’ of their natural realizers.⁶⁵ How are we supposed to interpret the relation between W and N_R ? If W supervenes on N_A , that is, if W is determined by N_A , and if N_R is caused exclusively by W , then how can we make sense of the idea that N_A does not cause N_R ? In other words, if we accept that N_A is sufficient for W , then how can N_A not be sufficient for N_R , given that W causes N_R ? It seems that to accept (2) would require us to deny supervenience.

⁶⁴ According to naturalism, if a natural event has a cause, the cause must be natural too. This is what it means for the natural world to be causally closed.

⁶⁵ I have already claimed in 1.4.3 that such a picture is implausible.

If we choose the third option, then we have a picture where N_R is causally overdetermined. It is not easy to explain how this kind of overdetermination is supposed to work though, because it looks like N_A does all of the causal work, leaving nothing for W to do. But if we are going to insist on (3), then our view will collapse into (2) or we must accept something like FitzPatrick's dual-aspect view. If we are going to say that W is necessary for N_R , then N_A alone cannot cause N_R , which could mean we have to accept that W floats free of N_A . But then we will have (2). Alternatively, we can say that N_R is caused both by N_A and W because N_A and W are two different aspects of the same entity. This is FitzPatrick's dual-aspect view. I have already argued in 1.4.3 that such a view is problematic due to its commitment to a bloated ontology and due to its failure to capture the *one-way* necessary dependence of moral properties on natural properties. Thus, (1) seems to be the most plausible option; that is, your reactions to morally wrong actions are caused exclusively by the natural properties associated with those actions.

The implausibility of (2) and (3) indicates that there is no good reason to posit moral properties because it is quite difficult, if not impossible, to fit such properties into the causal structure of the world. If moral properties have no causal power, then positing them would inflate our ontology for no clear reason at all. Note that I am not against postulating moral properties simply because they are mysterious. Rather, I support the claim that moral properties do not seem to play any role in accounting for our moral beliefs.⁶⁶ Even if we intuitively accept the existence of moral properties, the explanatory

⁶⁶ This objection is first put forward by Harman (1977, 6–9), who raises skepticism on the causal role of moral properties in explaining our moral beliefs. He compares two scenarios: (1) a scientist sees a vapor trail in a cloud chamber and believes that there is a proton in the chamber; (2) someone sees children setting a cat on fire and thinks what they do is evil. Harman argues that there is a sharp contrast between the cases (1) and (2). While the belief in (1) is explained by the fact that there is actually a proton in the chamber, there is no *moral* fact – in addition to the facts

impotence of such properties gives us reason that we should not incorporate them into our moral ontology.

The second reason to believe that moral intuitions are unreliable is the existence of moral disagreement. The objection may come in various forms. First, one could assert that intuitions are unreliable because sometimes there are irreconcilable conflicts among our intuitions. The idea is that if we base our moral beliefs on intuitions, then this may lead to unresolvable disagreements. It might be impossible to resolve a disagreement when two people have different and incompatible intuitions about a moral issue, and, the idea goes, this raises doubts on the reliability of moral intuitions. This, I believe, is not a strong argument because it does not attack intuitionism directly; rather, it only *predicts* what could follow from intuitionism. In fact, there *are*, and probably always will be, differing and conflicting intuitions about ethical matters, just as there are differing and conflicting intuitions about scientific, religious, or political matters. If it is unreasonable to cite disagreement as evidence of lack of truth or knowledge in these areas, then it is equally unreasonable to cite disagreement as evidence of lack of truth or knowledge in ethics, unless the argument is supported by evidence attacking the credibility of moral intuitions directly. The objection, in and of itself, does not speak against intuitionism.

One could also argue that the existence of disagreement undermines the idea that there are objective moral truths. That is, no moral judgment can be objectively true given that we don't *all* agree on moral issues. This objection is equally toothless because objectivity

constituting the action (setting the cat on fire) and psychological facts about the observer – that explains the belief in (2). Since (non-natural) moral properties and facts are irrelevant in explaining our moral beliefs, Harman concludes that we should not posit them. The above discussion bolsters Harman's objection because it does not only raise skepticism on the causal role of moral properties, but it also provides positive reason to refuse the idea that moral properties have causal powers.

does not entail agreement. You may disagree with someone about the moral status of female genital mutilation, for example, but such a practice may still be objectively wrong. It could just be that one of you is missing an objective moral truth due to wrong non-moral beliefs, due to certain practical pressures, or due to some other distorting factor. Objectivity rather entails that we do not create truths solely by forming beliefs; we may be wrong in our beliefs. People believe all sorts of odd things. Some people believe that the earth is flat, for instance, but their conviction does not transform the earth into a flat object. This is because it is an *objective* fact that the earth is round. The disagreement between flat earthers and the rest of the world does not undermine the objectivity of that fact. Likewise, one cannot undermine moral objectivity just by pointing to the differences in our moral beliefs.

Finally, one could assert that there is widespread disagreement about basic moral values but there is no such fundamental disagreement about sensory perceptions. If this is so, how can we know our intuitions are the right ones? I believe this is the most plausible form of the argument from disagreement because if people generally have different intuitions about basic moral values, then this means that people are often wrong about their moral intuitions. The existence of objective moral facts entails that whenever people's intuitions conflict, someone must be wrong. If intuitions vary greatly among individuals and cultures, then people often get things wrong. And since we cannot rationally assume that our intuitions are more accurate than those of others, moral intuitions are unreliable, and so we shouldn't trust them.

In response, one might say that certain basic moral values do not vary but are shared by all individuals and cultures. The reason people have differing intuitions about moral

issues could be that these universal values have been applied in different ways. More specifically, people's intuitions may vary due to their non-moral beliefs and due to different environmental circumstances. For example, people who think female genital mutilation is morally permissible might have this intuition because of their non-moral belief that excision is beneficial to society. In societies where infanticide is common, people could have been forced by circumstances to kill some of their babies even though they think they have moral obligations to care for their children. Scarcity of food, nomadic lifestyle, lack of birth control, harsh weather conditions, and the like, could force societies to adopt such practices, even though they do not lack fundamental respect for human life.⁶⁷ Admittedly, people disagree about moral issues such as abortion, euthanasia, or death penalty; however, the thought goes, people do *not* disagree on basic moral values and rules such as the value of looking after one's children, rules against murder, lying, stealing, and so on. You may disagree with someone over the moral status of animals, for example, but your intuition that "inflicting harm for mere amusement is wrong" is shared by all normally functioning human beings.

If this response is plausible, then intuitions may still be trustworthy. In fact, I find myself sympathetic to such a response. Obviously, our intuitions about abortion, euthanasia, animal rights, and so on, are controversial, which means that one might be wrong about their intuitions about such moral issues. Nevertheless, we also have *uncontroversial* moral intuitions such as "inflicting harm for mere amusement is wrong," "one should care for their children," "one should not steal," or "killing is normally wrong." We cannot say that widespread disagreement over moral issues render our

⁶⁷ Cf. Rachels 2019, 21–3.

intuitions unreliable because there seems to be *no* widespread disagreement over certain basic moral values and rules.

This response, however, is hasty. There really seems to be uncontroversial intuitions on which we base many of our moral beliefs. However, the important question here is, what is the *source* of these uncontroversial intuitions? Are we really perceiving non-natural moral properties of things when we have moral intuitions? Is this really the best explanation for our having those intuitions? Think about the basic moral rules that I mentioned. Rules against murder, lying, and stealing seem to be *necessary* for any society to exist at all. It is probable that we would end up with a Hobbesian state of nature if killing, lying, or stealing were not condemned. If lying were commonplace, then nobody would trust what anyone says, which would make communication impossible. If killing were commonplace, then nobody would feel safe, and everyone would do whatever necessary to avoid others. No society could have flourished under such circumstances. But if this is true, then we seem to have a more plausible, scientifically-backed explanation for our moral intuitions.⁶⁸ Since having certain basic moral intuitions is necessary for societies to exist and flourish, it could be that we have these intuitions due to evolutionary pressures. It is likely that a group who had certain moral intuitions had a greater chance of survival and reproductive success than other groups who lacked such intuitions. Instead of explaining moral beliefs by appeal to mysterious moral perception, we can explain them by appeal to evolutionary forces. But if our moral intuitions have been shaped by evolutionary mechanisms, those intuitions probably cannot give us access

⁶⁸ For such an account, see Street 2006, 127–30.

to objective, attitude-independent moral facts, if they exist at all. This is the idea behind evolutionary debunking arguments.

Evolutionary debunking arguments (EDAs) attempt to show that moral knowledge is improbable by placing a special focus on the evolutionary origins of our moral beliefs. The reasoning is as follows: according to moral realism, our moral beliefs are correct when they accurately represent objective (stance-independent) moral facts. However, it seems that “our system of evaluative judgments is thoroughly saturated by evolutionary influence.”⁶⁹ This means that our evaluative (and moral) intuitions – along with our evaluative (and moral) beliefs which are based on such intuitions – are simply a function of what enabled our hunter-gatherer ancestors to out-reproduce their peers and to maximize their reproductive output. But since realists do not take moral truth to be a function of what caused our Pleistocene ancestors to increase their reproductive success, it would be an “incredible coincidence” if our evolutionary shaped moral beliefs accurately represented stance-independent moral truths. Thus, given the evolutionary origins of our moral intuitions and the truth of realism, our moral intuitions are epistemically unreliable. The fact that the truth of moral realism leads to the epistemically unpleasant conclusion that we probably lack moral knowledge renders moral realism an implausible metaethical view.

There are *three* main types of EDAs. First, you can try to undermine the justification of all evaluative beliefs like Street (2006). Second, you can target only at our moral beliefs like Joyce (2006). (Note that the former kind of EDA is more ambitious than the latter since its target includes a greater number of beliefs.) Third, you can try to debunk a

⁶⁹ Street 2006, 114.

certain subset of moral beliefs to support a specific normative theory like Greene (2008) or de Lazari-Radek and Singer (2014). For example, you can raise skepticism on our deontological beliefs by arguing that they have been shaped by natural selection pressures unlike our consequentialist beliefs, which are the product of autonomous moral reasoning. So, some EDAs are more ambitious than the others. All EDAs, however, take a common form: they all claim that a certain subset of beliefs is unjustified, since (1) such beliefs have been shaped exclusively by evolutionary mechanisms and (2) evolutionary mechanisms aim at survival and reproductive success and thus are *insensitive* to stance-independent evaluative or moral truths, if there are any. (I will talk in detail about different versions of EDAs and about how a realist could respond to each version in the second chapter.) Moreover, differences between different versions of EDAs will become important when I argue against Vavova's (2014) claim that ambitious EDAs are self-refuting. But for now, it is important to see why EDAs constitute the most pressing epistemic objection to non-naturalism.

First, EDAs seem to give us a strong *inductive* reason to doubt the realist construal of moral perception. If there are stance-independent moral truths as realism asserts, and if there is a pervasive evolutionary influence on our moral intuitions, then our moral intuitions are *probably* not aligned with those independent moral truths, even if such truths exist. One reason for this could be that there seems to be no way of checking a particular moral intuition *independently* (i.e., without relying on other distorted moral intuitions) to see whether it is correct. There is a quick response to such reasoning. One could say that we cannot independently check on the reliability of our senses or memories either, so, by the same logic, we must dismiss them as unreliable too. A supporter of

phenomenal conservatism could say that the EDA strategy leads to a wholesale skepticism about *all* kinds of beliefs because there seems to be no difference between asking the realist to prove the reliability of our moral intuitions and asking someone to prove the reliability of their senses or memories. One may never be able to prove that we are *not* mistaken about our senses or memories due to some possible distorting factor. So, the idea goes, we should keep trusting our moral intuitions, just as we trust our senses and memories.

Such a response, however, is based on a misconception about EDAs. It could be true that a debunking argument would lead to a pervasive skepticism about *all* of our beliefs if it placed the burden of proof on the realist. But it does not. The debunker is concerned *only* with the reliability of our evaluative or moral intuitions, as described by realism. To this end, they give scientific evidence that gives us reason to think that our evaluative or moral intuitions, understood in a realist way, are probably mistaken due to the distorting influence of evolutionary forces on such intuitions. That is, the debunker is not asking the realist to prove the reliability of our intuitions; rather, they are giving us evidence of error (i.e., a reason that we are probably mistaken in our evaluative or moral intuitions) that follows from an empirical claim (i.e., evolutionary evidence). This enables the debunker to selectively claim that we cannot rationally form our beliefs on the basis of evaluative or moral intuitions, even if we concede that there are objective moral truths.⁷⁰ Thus, the burden of proof is on the debunker to show that there is really an extensive evolutionary influence on our intuitions and beliefs.⁷¹

⁷⁰ Cf. Vavova 2014, 82–7.

⁷¹ I explain in the second chapter (2.1.2) why the *scope* of evolutionary explanation of our moral intuitions and beliefs is of utmost importance in determining the strength of an EDA.

Moreover, EDAs do not assert that we *necessarily* are mistaken in our intuitions. Rather, they assert that we *probably* are mistaken in our intuitions, given our evolutionary and epistemic conditions. So, EDAs are *inductive* arguments.⁷² An inductive argument infers from a limited number of observations to a general, probabilistic conclusion. For instance, when we reach the conclusion that the sun will probably rise tomorrow from the set of observations that the sun has risen regularly so far, we give an inductive argument. The idea is that we have strong inductive reason to think that we are mistaken about our moral intuitions – provided that we have a *complete* evolutionary genealogy of our moral intuitions and beliefs – just as we have strong inductive reason to think that the sun will rise tomorrow. At no point does the debunker ask the realist to prove their intuitions to be trustworthy. The debunker is the one who provides reasons.

The second reason I take EDAs to be the most important *epistemic* argument against non-naturalism is that other important epistemic objections ultimately rest on *metaphysical* considerations. Harman's argument from explanatory impotence, as well as other objections that raise skepticism on the causal role of moral properties in explaining our moral beliefs, get their force from one of the least controversial theses in metaethics, namely the supervenience thesis. Those objections are definitely worth considering as serious epistemic challenges because the metaphysical nature of moral properties has significant epistemic implications: since the realist cannot plausibly account for supervenience, it is hard to make sense of how moral properties are supposed to cause our intuitions. Thus, their perceptual characterization of moral knowledge seems to fail. Conversely, EDAs are not metaphysically motivated arguments since they remain

⁷² For more on this, see 2.1.2.

agnostic about the existence of moral properties. This is because the argument is a *conditional*: if there are stance-independent moral properties and facts, we *probably* cannot know them due to our evolutionary history and epistemic position. The skeptical conclusion of an EDA holds regardless of the existence of objective moral properties and facts. Hence, EDAs are purely epistemic arguments.

Some think that EDAs make controversial metaphysical claims about the nature of morality. For example, Ramon Das (2015) contends that all EDAs share a hidden metaphysical assumption that moral facts are *irreducible* to natural facts. According to him, to debunk evaluative or moral beliefs, it is not enough simply to provide a causal explanation of their origins that does not invoke their truth. That would prove too much, i.e., that would lead to a “wholesale debunking of beliefs regarding any common-sense object or property that could be given a reductive explanation that did not appeal to truth of the relevant macro-level entity” (Das 2015, 422). Das thinks that if we assume from the outset that the entity or phenomenon in question cannot be described naturalistically, then we can debunk almost anything by simply giving a causal explanation of how something is perceived or felt. This would apply not only to everyday objects we perceive but also to our mental states. Das claims that the anti-naturalistic position of the debunker follows from their prior rejection of objective moral values. Otherwise, naturalism would be a viable option for them. His idea is that since EDAs quickly dismiss objective moral values as unjustified rather than providing a naturalistic account of those values that could escape the debunking argument, it seems that the debunker is committed to the claim that there are no objective values from the outset.

As I mentioned above, EDAs do not lead to a wholesale skepticism about all kinds of beliefs because they do not place the burden of proof on realism; rather, they give us reason to believe that we are probably mistaken about our moral intuitions and beliefs. Furthermore, EDAs take the form of a *reductio*. They first assume the truth of realism and the existence of objective values. They then assert that we cannot rationally hold such a view due to the evolutionary explanation of our moral intuitions and beliefs. This does not entail that EDAs reject outright the existence of objective values. EDAs do not reject naturalism or reductionism either. They merely contend that the truth of realism leaves us with an epistemically unappealing picture: we cannot justify our moral beliefs because they can be given a complete evolutionary explanation. The target here is the *stance-independence* condition of realism. As Street (2006, 136) puts it, when it comes to whether naturalism can escape the evolutionary challenge, the crucial question is, “does the view in question understand evaluative truths as holding, in a fully robust way, independently of all our evaluative attitudes?” (The question is not, does the view in question understand evaluative or moral properties as being reducible to natural properties?) If which natural properties moral properties are identical with *depends* on our contingent attitudes, as Railton’s (1986) naturalistic reduction in terms of an ideal response has it, then the view can escape the epistemic challenge posed by EDAs.⁷³ And

⁷³ As I described in 1.4.2, Railton identifies moral goodness with what is non-morally good for the agent, and people’s non-moral good is determined by their constitution and circumstances. It follows from this that what moral goodness is identical with is relative to the agent’s contingent attitudes. Suppose that A’s ideal counterpart, A+, wants A to want to care about her well-being in circumstances C. In different circumstances, namely D, A+’s attitudes could change such that A+ would no longer want A to want to care about her well-being. Note that facts about natural-moral identities are dependent upon our contingent attitudes under Railton’s view, because (a) what we would desire to desire under conditions of full information and rationality is subject to change, and (b) moral facts are identical to what our attitudes would pick out as good under certain circumstances (cf. Street 2006, 137).

if a moral naturalist holds the view that facts about natural-moral identities are *independent* of our attitudes, then their view is subject to the evolutionary challenge. So, EDAs attack realism rather than reductive naturalism.⁷⁴ Once again, EDAs are purely epistemic arguments.

Finally, evolutionary explanations of our moral intuitions and beliefs may pose a problem for *all* kinds of non-naturalism. There is a possible non-naturalist solution to other objections that I mentioned. For example, the problem of explaining supervenience could in principle be avoided by adopting a Parfitian moral ontology complemented with Kantian constitutivism. We don't need to explain why there is a one-way necessary dependence of moral properties on natural properties if we maintain that moral properties and facts exist in a non-ontological way. Moreover, epistemic problems that originate from the causal inertness and explanatory impotence of moral properties can be avoided in a similar manner. The fact that moral properties and facts are causally inert poses a serious problem for the realist construal of moral knowledge because it is far from obvious how moral perception, that is, interaction between moral properties and our minds, is supposed to occur. But if moral principles, along with the principles of logic and mathematics, are not 'out there' to be discovered (or if they are not tied to human rationality as an absolute inner value) but are constitutive of how our reason necessarily functions, it may become easier for us to explain how we gain moral knowledge.⁷⁵

Nevertheless, I don't think we can circumvent the epistemic challenge posed by EDAs

⁷⁴ Different versions of EDAs attack different kinds of realism. For example, Street's (2006) EDA attacks evaluative realism, while Joyce's (2006) EDA attacks moral realism.

⁷⁵ For more on this, see 3.4.6.

easily just by adopting this non-ontological option. In fact, the idea that moral intuitions are determined exclusively by evolutionary processes seems to be a fundamental threat to assumptions that underlie this dissertation.

What would make the non-ontological non-naturalism a plausible alternative to moral realism? Why do we need an alternative to moral realism in the first place? Why do I reject naturalism? The answers to these questions reveal the two main assumptions that underlie this dissertation. First, non-ontological non-naturalism can possibly be a plausible alternative to moral realism if we complement it with a Kantian constitutivist origins account, because many problems associated with the non-naturalist versions of realism arise from the stance-independence condition. The hope is that we can plausibly talk about *stance-dependent objective* moral truths and thereby present a less problematic account of morality than realism that can also retain morality's objectivity. So, the first assumption is that *morality, if it exists, is objective*. Second, naturalism is not the solution I am looking for because even though naturalistic accounts could be plausible in many respects, they fail to capture categorical normativity, which is an indispensable feature of morality. Here is the second assumption: *morality, if it exists, is categorically normative*. These two main assumptions constitute the starting point of my inquiry and inform my further investigations.

What if our senses of objectivity and categorical normativity themselves are the product of evolutionary processes? Perhaps we think morality is objective and categorically normative not because morality really is so but because having such beliefs or intuitions makes us good social cooperators. Richard Joyce (2001, 141–8) offers a compelling story of how evolutionary forces could have shaped our sense of 'moral

ought' and our capacity for normative guidance: altruistic behavior towards strangers promotes cooperation and thereby increases survival and reproductive success. However, someone who accepts help but does not reciprocate is a big threat to the group. Thus, it is likely that groups who were sensitive and hostile towards exploiters had a greater chance to flourish compared to groups who were not. Moreover, the trait of sensitivity and hostility towards exploiters works best if it is accompanied by a sensitivity to others' *motivations*. The motivation of a companion is important because you cannot really trust a person who helps you with the expectation of getting help back. Such calculating cooperators are fickle; they will abandon you once they think things are not working to their advantage. You would certainly want a companion who likes you and cares about you. But you cannot rely solely on sympathy either because a life-threatening situation may dissipate compassionate feelings. Rather, the appropriate motivation that accompanies cooperative behavior is a sense of 'ought,' namely a feeling that certain acts are obligatory. You would prefer a companion who has a feeling of obligation to help you.⁷⁶ Of course, sympathy accompanied by a feeling of obligation would be optimally efficient for ensuring cooperative behavior. Indeed, cooperative behavior can also be regulated by expectation of self-benefit. But cooperative behavior is *best* regulated by "the strongest sense of authority available: the requirement that must be obeyed regardless" (Joyce 2001, 145).

According to Joyce (2001, 146–7), our sense of categorical normativity and our capacity for normative guidance are *innate*. That is, the disposition to believe in the existence of categorically normative moral requirements is deeply ingrained in our

⁷⁶ What you would 'want' or 'prefer,' in this context, is merely a shorthand for what kind of motivation would best promote cooperation.

psychological makeup and is deeply embedded in our genetic code. We also have a psychological disposition to believe that certain kinds of action, such as looking after one's children, keeping promises, repaying debts, and so on, are morally required. Admittedly, these dispositions become manifest only when an individual interacts with other people. A person who is isolated from others will not believe that there are moral requirements. However, we all have these psychological dispositions in appropriate circumstances.

Joyce's evolutionary story threatens to undermine the assumption that morality is objective and categorically normative because if it is true, then morality is normative only in an instrumental or aim-relative way: *if* you are to survive and reproduce, you must cooperate with others as efficiently as possible, and the best way to do that is to have a feeling of moral obligation. Moreover, Joyce's story indicates that morality is based on our *feelings*. Feelings are contingent and relative to the circumstances. They differ from person to person as well as from group to group. But morality is objective only if it is independent of our desires, feelings, or inclinations. Thus, feelings cannot ground objective morality.⁷⁷ Furthermore, this evolutionary picture, at least on the face of it, is quite different from Kantian constitutivism, according to which moral principles are constitutive of how our reason *necessarily* functions.⁷⁸ That is, if the constitutivist story is

⁷⁷ For example, Kant writes, “[f]rom the feeling of a sensation that may be different in every creature, no generally valid law can be derived for all thinking beings, and that is how the moral principle must be constituted” (VE 29:625).

⁷⁸ Oliver Sensen (2013, 2017) makes a compelling case that Kant embraces this kind of constitutivism in his moral philosophy. However, my aim is not to discuss whether Kant himself is a constitutivist. Rather, my aim is to investigate whether such a view could be presented as a plausible alternative to moral realism.

true, then our reason, independently of any condition, gives us moral laws that apply to all rational beings. This means that we are all bound by moral requirements independently of our evolutionary history.

The idea of an innate moral sense seems to be at odds with Kantian constitutivism because if our moral sense is implanted in us by evolutionary processes, then morality cannot be necessary and universal: we could have had a different moral sense under different circumstances. According to Kant, an innate moral sense would only yield a “subjective necessity, arbitrarily implanted in us” (KrV B168). This means that moral principles would not be valid for all rational beings, but it would be valid only for beings who have an evolutionary history similar to ours.⁷⁹ But in Kant’s view, “we must not let ourselves think of wanting to derive the reality of this [moral] principle from the *special property of human nature*” (GMS 425).⁸⁰ That is, a necessary and universal morality cannot be based on evolved human nature. Instead, moral obligation “is to be practical unconditional necessity of action and it must therefore hold for all rational beings” (ibid.).

At first glance, the truth of Joyce’s evolutionary explanation could have devastating implications for my investigation. If our sense of moral objectivity and categorical normativity is merely the product of human evolution, then it seems that I cannot reject naturalism on the grounds that it fails to capture the categorical authority of moral norms.

⁷⁹ Cf. Sensen 2013, 76.

⁸⁰ Kant here does not talk about evolution. Rather, this claim is part of his argument that *a priori* elements of our cognition cannot be necessary and universal if we conceive of them as being implanted in us by God. However, this idea can be applied to evolution as well. On this point, see Sensen 2013, 75–6; 2017, 202. More on this in 3.4.4.

Also, searching for an alternative to moral realism would be a futile endeavor, since the truth of the evolutionary story indicates that morality is not objective after all. However, we must be cautious here, as the evolutionary story does not give *decisive* evidence that moral objectivity and categorical normativity are merely illusions. Recall that EDAs are essentially inductive arguments with probabilistic conclusions. For all the talk of evolutionary influence on our moral intuitions and beliefs, morality could still be objective and categorically normative. But if we have the sense of objectivity and categorical normativity simply because of evolutionary pressures, it would be an enormous coincidence if morality turned out to be exactly as expected, i.e., if our evolutionarily shaped moral sense got things right. So, if the evolutionary explanation is correct, we seem to have strong inductive reason to believe that morality lacks objectivity and normativity. Thus, a plausible alternative to moral realism must address this pressing epistemic problem, either by offering an uncompromising account of objectivity and normativity that is also compatible with the evolutionary story, or by showing that the evolutionary challenge is not sufficient to undermine the alternative view (or both). I will discuss this issue in the following chapters.

CHAPTER 2: EVOLUTIONARY DEBUNKING ARGUMENTS

Evolutionary explanations of our moral beliefs or intuitions may pose a serious problem not only for moral realism but also for all kinds of non-naturalism, including the version of Kantian constitutivism that I defend in the third chapter. As I mentioned in the first chapter, the idea that moral beliefs or intuitions are determined exclusively by evolutionary processes seems to be a fundamental threat to two main assumptions that underline this dissertation: (1) morality, if it exists, is objective; and (2) morality, if it exists, is categorically normative. If it is true that we think morality is objective and categorically normative not because morality really is so but because having such beliefs or intuitions makes us good social cooperators, then both realism and Kantian constitutivism could be epistemically implausible. Even if morality is objective and categorically normative, how do we know that? Realism talks about stance-independent moral properties and facts, whereas Kantian constitutivism bases objectivity and categorical normativity on the standpoint of reason – morality is *a priori* prescribed by our reason. Both stance-independent properties and the standpoint of reason are independent of the aim of survival and reproduction. And the question is, how do we know that morality has this *a priori* status when our moral beliefs and intuitions are exclusively shaped by evolutionary forces? Thus, EDAs, *prima facie*, pose a serious epistemic threat to non-naturalist accounts of moral objectivity.

However, (a) EDAs are not strong enough to undermine either position, and (b) Kantian constitutivism is compatible with the idea that our sense of moral objectivity and categorical normativity have increased our survival and reproductive success. I defend (a) in this chapter, and (b) in the third chapter. To defend (a), I first claim that the ambition of an EDA affects the argument's empirical premise: the more set of beliefs an EDA calls into question the more difficult it becomes to provide a complete evolutionary origins story (2.1). I then discuss why evolutionary explanations are insufficient for some of our moral beliefs and intuitions, i.e., why the empirical premise of an EDA is not as strong as the debunker believes it is (2.2). I talk about different types of moral intuitions and claim that our theoretical and formal moral intuitions are immune to direct evolutionary influence. I argue that the processes of autonomous (gene-independent) moral reasoning and cultural evolution together make it possible for us to arrive at evaluative/moral judgments independent of selective pressures. I primarily focus on Joshua Greene's (2008) EDA and reveal the weakness of its empirical premise. Since my argument against Greene's EDA works also against other kinds of EDA, I conclude that EDAs fail to pose a strong epistemic challenge for evaluative/moral realism and for other non-naturalist accounts of moral objectivity.

2.1 How to Respond to Evolutionary Debunking Arguments

My aim in this section is to show how to plausibly respond to evolutionary debunking arguments. To do that, I first lay out the epistemic challenge EDAs pose for evaluative/moral realism (2.1.1) and briefly explain Sharon Street's (2.1.1.2) and Richard Joyce's arguments (2.1.1.3). I then discuss Katia Vavova's objection that ambitious

EDAs are self-refuting (2.1.2). I argue, contra Vavova, that the level of ambition of an EDA does not affect the strength of its epistemic premise because EDAs are essentially *inductive* arguments. Rather, the level of ambition of an EDA affects the strength of its *empirical* premise, which is the Achilles heel of any ambitious EDA. Finally, I respond to a possible objection that Vavova's point is a *dialectical* one concerning the epistemology of disagreement and the possibility of a debunking being successful, rather than a *logical* one about the internal structure of an argument (2.1.3).

2.1.1 Evolutionary Challenge

2.1.1.1 The Structure of an Evolutionary Debunking Argument

Evolutionary debunking arguments claim to undermine the justification of our evaluative beliefs by placing a special focus on the evolutionary origins of them. Some of such arguments are more ambitious than the others as they try to undermine the justification of *all* evaluative beliefs (Street 2006), while some of them are targeted at moral beliefs only (Joyce 2006), and some at a certain subset of moral beliefs (Greene 2008). All EDAs, however, take a common form. They all claim that knowledge of a certain subset of evaluative beliefs is improbable, since (i) such beliefs are shaped exclusively by the mechanisms of natural selection and (ii) evolutionary processes aim at survival and reproductive success and thus are insensitive to attitude-independent evaluative truths, if there are any. The former is the *empirical* premise, and the latter is the *insensitivity* premise. EDAs also have an *epistemic* premise, namely that if non-naturalist evaluative (or moral) realism, the empirical premise, and the insensitivity premise are true, then we

cannot justify the beliefs in question. These three premises constitute the blueprint of any EDA:

- (1) *Empirical premise.* Evolutionary mechanisms have a pervasive influence on the content of our evaluative/moral beliefs.
- (2) *Insensitivity premise.* Evolutionary mechanisms aim at survival and reproductive success and not attitude-independent evaluative/moral truths.
- (3) *Epistemic premise.* If there are attitude-independent evaluative/moral truths, evolutionary mechanisms have a pervasive influence on the content of our evaluative/moral beliefs, and evolutionary mechanisms aim at survival and reproductive success and not attitude-independent evaluative/moral truths, then we lack an independent reason to think that our evaluative/moral beliefs track the truth, i.e., we lack justification for our evaluative/moral beliefs.
- (4) *Skeptical conclusion.* We lack knowledge of attitude-independent evaluative/moral truths, if they exist at all.

The epistemic premise is the core of any debunking argument. There are many ways of forming beliefs. Think, for example, of people who rest their beliefs about an outcome of a football game, or about whether it is going to rain the following day on the behavior of animal oracles. Although their beliefs might turn out to be true, they are only incidentally true since animal behavior has nothing to do with the states of affairs in a football game or with the state of the atmosphere. Hence, we have a good reason to suppose that people who form their beliefs through a process that is not good at tracking the truth – like in the case of animal oracles – are not justified in their beliefs. Kahane (2011, 106) calls such processes ‘off-track processes.’ Hearing, on the other hand, is most of the time an epistemically reliable process, which means that it is good at tracking the truth. People whose beliefs are informed by their hearing mechanism are correct in their beliefs about what they hear, provided that they do not have an impaired hearing mechanism and no environmental factor is distorting their beliefs. For example, the fact that I hear Beethoven’s Eighth Symphony on the radio at the moment is a good

justification for my belief that Beethoven's Eighth Symphony is now playing in the radio, provided that there is nothing unusual about my hearing mechanism and nothing in my surroundings affects its proper function. EDAs, therefore, are based on the crucial distinction between processes that track the truth and off-track processes.

2.1.1.2 Street

According to Sharon Street's empirical premise, evolutionary processes have an enormous influence on the content of our evaluative beliefs. Although evolutionary forces do not *directly* determine the content of our evaluative beliefs, they select for "basic behavioral and motivational tendencies" (Street 2006, 113), which in turn play an important role in shaping the content of these beliefs. This is the reason why there is a noticeable match between some of the basic evaluative judgments made by different cultures in different times. Street mentions six of them: (1) to promote one's survival; (2) to promote our children's survival; (3) to promote our family members' survival; (4) to reciprocate when treated well; (5) to praise and reward altruistic behavior; (6) to blame and punish actions involving deliberate harm (Street 2006, 115). In Street's view, the explanation of why we share these basic evaluative beliefs lies in the fact that people who have psychological tendencies to form these basic beliefs tend to survive and reproduce in more numbers than those who lack such tendencies. Moreover, Street asserts that there is a conspicuous similarity between many of our widely held evaluative beliefs and basic evaluative tendencies of animals, which counts in favor of the empirical premise.¹

¹ We can observe many examples of survival promotion, reciprocal and altruistic behavior across a variety of animal species (Street 2006, 117). There are a number of studies that verify Street's claim. For example, according to Frans de Waal's (2014) study, evolved behavior of nonhuman

After giving the empirical premise, Street presents her Darwinian Dilemma: either there is a relation between evolutionary influences on our evaluative beliefs and independent moral truths or there is not. If there is no causal connection between evolutionary mechanisms and moral truths, then evolutionary forces probably have a distorting influence on our evaluative beliefs, which are off-track. Street claims that there are infinitely many logically possible coherent belief systems and that according to the realist one of them is the right one (Street 2006, 122). And since our belief system is mostly determined by selective pressures and not by independent evaluative truths, it would be an enormous and inexplicable coincidence if our belief system turned out to be the right one. Thus, she concludes that we are most likely to be wrong in our evaluative beliefs.

Street then explains why assuming a relation between selective pressures and evaluative truths is not a good strategy for the realist either. The realist, of course, must explain what kind of relation holds between these two. Street argues that the only option the realist has is to adopt a *tracking account*, according to which selective pressures aim at evaluative truths because true evaluative beliefs are fitness-enhancing.² Tracking account, in Street's view, is a scientific account that explains the way evolutionary mechanisms operate and gives a reason as to why certain basic evaluative beliefs are

animals involves normativity as "adherence to an ideal or standard" because they pursue social values like humans. They actively try to maintain cooperation and harmony by "reconciling after conflict, protesting against unequal divisions, and breaking up fights amongst others." Doing so allows them to correct deviations from an ideal state (de Waal 2014, 200).

² One example is Derek Parfit's (2011a, 21–2) tracking account: "Just as cheetahs were selected for their speed, and giraffes were selected for their long necks, human beings were selected for their rationality, which chiefly consists in their ability to respond to reasons."

chosen in most cultures over others. Thus, it competes with other scientific accounts (Street 2006, 126). Street argues that tracking account loses the scientific battle against what she calls the *adaptive link account* in terms of scientific parsimony and clarity. The adaptive link account claims that our tendencies to prefer certain evaluative beliefs over others are evolutionarily advantageous because they motivate us to respond environmental conditions in fitness-enhancing ways (Street 2006, 127).³ The adaptive link account is more parsimonious than realism, for unlike realism it does not posit something additional, namely the existence of independent evaluative truths. It is also clearer than the realist account, for realism explains the evolutionary benefit of certain evaluative beliefs only by appealing to the fact that *those beliefs are true*. In other words, moral realism does not say anything about *why* those beliefs promote survival. This makes the realist account obscure and question-begging (Street 2006, 129–30).

2.1.1.3 Joyce

Joyce's EDA aims exclusively at our moral beliefs. Like Street, he believes that we have a tendency to make certain moral judgments and that that tendency is an adaptation (Joyce 2006, ch.4). That is, he believes that we have an innate moral sense that produces similar basic moral judgments, and that we acquired this moral sense because it is biologically advantageous.⁴ Joyce also adopts a realist semantics as regards moral

³ For example, we tend to have evaluative beliefs that encourage cooperation because the average human life would considerably be shortened if cheating were the norm.

⁴ Although Joyce believes that biology determines the content of moral beliefs to the extent that it reveals "universal characteristics of morality" such as the wrongness of killing, he maintains that the differences in our moral judgments are mainly because of the effects of cultural transmission. However, he thinks that the influence of culture on the content of our moral beliefs is consistent

propositions, following the lead of J. L. Mackie (1977): he believes that moral judgments are rationally authoritative by virtue of their nature, i.e., morality inescapably provides us with attitude-independent (or categorical) reasons (Joyce 2006, 192).

Joyce's empirical premise is that we have a complete and "confirmed non-moral genealogy" (Joyce 2006, 190). The idea is that we think there are natural, non-natural, or divinely given moral values that provide standards for how we should live. But evolutionary biology tells us that these standards have arisen simply because they facilitate cooperation among humans, which makes it easier for us to survive. Thus, "[w]ere it not for a certain social ancestry affecting our biology [...] we wouldn't have concepts like *obligation*, *virtue*, *property*, *desert*, and *fairness* at all" (Joyce 2006, 181). Joyce's insensitivity premise claims that the best explanation of our basic moral judgments is that "they are expressions of underlying 'design features' of human psychology" (Joyce 2006, 140). Just as we cannot justify our beliefs about the battle of Waterloo if they are caused by an imaginary belief pill that is insensitive to the facts about Waterloo, so we cannot justify our moral beliefs if they are generated by biological processes that are insensitive to proposed moral facts or truths. And Joyce's conclusion is that "we have no reason to believe in moral facts" (Joyce 2006, 210).

The important difference between Joyce's and Street's respective EDAs is that while Joyce thinks his EDA supports moral skepticism, Street thinks her EDA ultimately supports the truth of anti-realism. Joyce adopts a realist semantics *only* about moral propositions – that is, only moral propositions refer to categorical reasons. Street, on the other hand, believes that *all* evaluative discourse should be viewed in realist terms. Since

with an innate moral sense, which makes cultural transmission possible in the first place (Joyce 2006, 140).

realism itself is an evaluative claim, we “must be to adjust our metaethical view so as to become antirealists” (Street 2006, 141). Thus, Street’s EDA implies not only moral skepticism but also complete evaluative skepticism.

2.1.2 Ambition and Strength

It is important to note that if the debunking argument places the burden of proof on the realist, the argument then collapses into a pervasive skepticism about *all* of our beliefs.⁵ The debunker’s aim is to undermine a limited set of beliefs using scientific evidence. But if the debunker’s argument asks the realist to provide an independent reason to think that their beliefs are not mistaken, then the empirical premise becomes superfluous, and the argument’s target extends to our entire body of belief. We may never provide good (independent) reasons for the truth of *any* of our beliefs due to *some* possible distorting factor, but this general epistemic worry has never been the debunker’s concern. Rather, the debunker is concerned with the rationality of our evaluative or moral beliefs. To this end, she gives scientific evidence that gives us reason to think that our evaluative or moral beliefs are mistaken due to the distorting influence of evolutionary forces on such beliefs. Thus, the burden of proof must be on the debunker: she needs to give us evidence of error (a good reason that we are probably mistaken in our evaluative or moral beliefs) that follows from an empirical claim (evolutionary evidence), rather than asking the realist to give independent reason that our beliefs are not mistaken. Only then she can selectively claim that we cannot rationally maintain our evaluative or moral beliefs.

⁵ I agree with Vavova (2014, 82–4) on this point.

When formulating the empirical premise, the debunker determines the target of her argument: empirical evidence could show that *all* of our *evaluative* beliefs have been shaped by evolutionary processes, or it could show that *all* of our *moral* beliefs are determined by such processes. Street's ambitious EDA chooses the former target, while Joyce's less ambitious EDA chooses the latter. On Vavova's account, Street's argument is less likely to succeed than Joyce's argument due to the epistemic principle that she calls "the Inverse Rule of Debunking." However, she thinks that both accounts are ambitious enough to fail.

2.1.2.1 The Inverse Rule of Debunking

Vavova's claim is that the more ambitious a debunking argument becomes the less prospect of success it has. This is because the debunker tries to give us good reasons to think that we are mistaken about a certain body of beliefs, and what makes a reason good is its independence from the set of beliefs that are called into doubt. For example, if the aim of your argument is to undermine all perceptual beliefs, it would become illegitimate to base any of your premises on the truth of your beliefs that are formed through your senses. Street's EDA calls all of our evaluative beliefs into doubt; thus, the independent ground that reveals our mistake in our evaluative beliefs cannot involve any evaluative claim. Joyce's EDA can employ our nonmoral evaluative beliefs as an independent ground for the evidence of error, since its target encompasses all of our moral beliefs. Vavova calls this epistemic principle "The Inverse Rule of Debunking," according to which "[t]he potential strength of a debunking argument is inversely proportional to its ambition" (Vavova 2014, 98).

Although I do not see any compelling reason to reject the principle, I argue that ambition of an EDA does not really affect the strength of its epistemic premise. Rather, the relation between ambition and strength of an EDA becomes relevant only with respect to the scope of evolutionary influence on the content of our beliefs, since EDAs are essentially inductive arguments. The right strategy to debunk the debunker is thus to attack the *empirical* premise of her argument, which determines the inductive strength of any EDA.

2.1.2.2 Self-Refutation Argument

Since the debunker's aim is not to deductively prove that our evaluative or moral beliefs are *necessarily* wrong but merely to show that such beliefs are *probably* wrong given our evolutionary and epistemic conditions, EDAs are *inductive* arguments. An inductive argument infers from a limited number of observations to a general, probabilistic conclusion. For instance, when we reach the conclusion that the sun will probably rise tomorrow from the set of observations that sun has risen regularly so far, we give an inductive argument. Similarly, an EDA reaches the conclusion that our evaluative or moral beliefs are *probably* wrong from an evolutionary explanation of such beliefs. When we are given an EDA, we realize that there is a discrepancy between what we (or realists, to be more specific) take evaluative/moral judgments to be, and how evolutionary psychology describes them. That is, we realize that the fact that evaluative/moral claims are based on evolved psychological dispositions that favor adaptive behaviors is at odds with the presupposed categorical (desire-independent) nature of morality. Then we infer that our evaluative/moral beliefs are probably mistaken, just as we infer that the sun will

probably rise tomorrow. Making such inferences is just one of the things our minds naturally do.

When we infer the probabilistic conclusion of an EDA from the scientific evidence it provides, we make an assumption about epistemic reasons, namely that scientific evidence has the power to undermine our intuitions. But what makes us believe that scientific evidence is epistemically more reliable than our intuitions? It is perfectly possible that having this evaluative belief is also an adaptation. Wouldn't this then threaten the kind of EDA that calls *all* of our evaluative beliefs into doubt? This is Vavova's objection to Street's EDA (2014, 87–9). Vavova argues that Street's EDA targets both practical and epistemic reasons, both of which have been shaped by natural selection. The idea is that if we cannot trust any of our evaluative beliefs, then we cannot trust our beliefs about whether our evaluative beliefs are debunked by the argument. “[T]o evaluate we must rely on the evaluative” (Vavova 2014, 89); however, if the argument aims to undermine *all* evaluative judgments, then we lack the resources to determine whether the targeted beliefs are debunked. Hence, Vavova concludes, Street's EDA is self-refuting.

I do not think Vavova's strategy delivers a knockout blow to Street's argument. Recall that an EDA is an inductive argument and that inductive arguments allow for their conclusion to be false due to their probabilistic nature. Street does not claim that our evaluative beliefs are *necessarily* wrong but that they are *probably* wrong given our evolutionary and epistemic conditions. The conclusion of Street's EDA allows the possibility that some of our evaluative beliefs turn out to be true, and it follows that these

true evaluative beliefs could include some of our beliefs about epistemic reasons, science, mathematics, and so on.

As a matter of fact, our beliefs about epistemic reasons are more likely to be true compared to our beliefs about concrete evaluative matters, even though the former are also beliefs. Our beliefs about epistemic reasons are *beliefs about beliefs* because they are about whether our beliefs are epistemically trustworthy. But beliefs about beliefs are categorically different from *beliefs about specific cases*. Notice the difference between the following two statements: (1) “Our evaluative beliefs are probably false unless supported by evidence;” (2) “My partner is beautiful.” The former statement is more likely to be a product of *reasoning* rather than *biological conditioning*, and thus it is more likely to be true. This is because the reasoning that is involved in the former statement forces one to distance themselves from their (possibly distorted) beliefs about specific cases and make them realize that such beliefs are epistemically vulnerable. And this process allows room for belief revision.

It is difficult to declare an EDA to be self-refuting unless its epistemic premise renders our beliefs about epistemic reasons wrong. I am pointing to the difference between the statements “All *Xs* are wrong” and “All *Xs* are *probably* wrong,” when both statements are themselves instances of *X*. It is more difficult to call the latter self-refuting because it leaves open the possibility of itself (or any other instance of *X*) being true. So, it is possible to take the statement “All of our evaluative beliefs are *probably* false, including this one” to show ultimately that *some* of our evaluative beliefs are true.

What about Joyce’s EDA? According to Vavova, EDAs that target moral beliefs, such as Joyce’s EDA, still target too much (2014, 90–3). Such EDAs claim that we are

probably mistaken about morality because our moral beliefs have been shaped by natural selection, which is an off-track process. However, Vavova argues, to show that there is no relation between adaptive moral beliefs and true moral beliefs, we first need to know something about the contents of true moral beliefs and adaptive moral beliefs. And this requires us to make assumptions about what morality is like. Otherwise, morality could be about anything and accordingly we would have no reason to think that adaptive moral beliefs and true moral beliefs do not coincide. EDAs like Joyce's make such assumptions (e.g., moral judgments are rationally authoritative) but they at the same time call our entire body of moral beliefs into doubt. They thereby render their own moral assumption illegitimate. Thus, EDAs that declare *all* of our moral beliefs to be epistemically suspect cannot give us independent reason to think that we are mistaken about morality.

I do not think such an objection refutes Joyce's EDA. Since the epistemic premise of an EDA takes the form of a *reductio*, it is essential for any EDA to assume something about morality or about normative domain in general. Otherwise, it would not go through. If your aim is to debunk moral realism, you should first assume that moral realism is true. If your aim is to debunk our moral beliefs altogether, you should first make an assumption about basic commitments and presuppositions to morality. The epistemic premise is a conditional: *If* our moral beliefs have such-and-such features, then they are *probably* wrong, considering the extensive influence of evolutionary forces on the content of those beliefs and the insensitivity of evolutionary processes to their truth.

Moral beliefs could, of course, have different features than it is assumed by the argument. It is possible that the correct account of morality is an *anti-realist* one. In that case, only the assumed conception of morality, that is moral realism, could be debunked.

This is where Joyce seems to go wrong. He takes the conclusion of his argument to have debunked morality in general. However, he thereby dismisses anti-realist conceptions of morality that could escape the evolutionary challenge.⁶ Although there seems to be nothing wrong with making a metaethical assumption to get the argument going, dismissing alternative conceptions of morality could possibly create a problem for the debunker, if their intention is to debunk morality as a whole.

2.1.2.3 Ambition and Inductive Strength

I have argued that the level of ambition of an EDA does not affect the strength of its epistemic premise because EDAs are essentially inductive arguments. Focusing on the epistemic premise and declaring more ambitious EDAs to be self-defeating do not remove the skeptical worry that we might be mistaken in our evaluative/moral beliefs. As long as one admits that our beliefs are heavily shaped by natural selection, it is natural and plausible to think that objective morality could simply be an illusion. This worry remains even if we think the argument is self-defeating.

Does this mean there is no relationship whatsoever between an EDA's ambition and strength? There *is* such a relationship, but the level of ambition only affects the strength of the *empirical* premise. The only way to ease the skeptical worry seems to be to show that there is *no extensive* evolutionary influence on our beliefs. Many philosophers and evolutionary biologists agree that certain capacities and tendencies relevant to evaluative thought and behavior, and *some* of the content of our evaluative beliefs *can* be explained

⁶ Street (2006, 152–4) does not fall into this trap and acknowledges that anti-realist conceptions of morality are safe against her EDA.

by evolution. However, there is much less agreement among them on whether evolutionary forces have a *pervasive* influence on the content of our evaluative beliefs. The idea is that the effects of *human culture* and *moral reasoning* on the contents of our evaluative beliefs can be thought of independently from the effects of biological evolution on such beliefs.⁷

If the debunker can show conclusively that her empirical premise is true, her argument will get very strong. However, the more sets of beliefs she claims to have determined by our biological nature, the more difficult it gets to provide a complete evolutionary origins story. For instance, it would be more difficult for a debunker who tries to debunk all of our evaluative beliefs to prove her empirical premise than a debunker who aims only at our moral beliefs. But provided that both debunkers succeed in their respective tasks, the former debunker's argument would get inductively stronger than the latter debunker's argument, for her empirical premise would encompass a greater number of beliefs. Thus, the level of ambition of an EDA has an effect on the strength of its empirical premise, and accordingly on the inductive strength of the argument.

2.1.3 Logical Property vs. Dialectical Property

I would like to address a possible concern about my rejection of Vavova's view. One could claim that it is a straw man for me to talk about a logical contradiction. After all, Vavova never uses the terms "self-refuting," "self-defeating," or "contradiction." Rather, she uses terms like "good reason," "independent reason," and "setting aside." Vavova

⁷ I will discuss this idea in the next section.

thinks, the objection goes, when someone presents you with purported reasons to doubt some of your beliefs, this “requires you to set aside the targeted beliefs when evaluating her challenge” (Vavova 2014, 98). And as the class of targeted beliefs grows in size, there are less “resources for both presenting and evaluating evidence of error” (ibid.). To evaluate Street’s EDA, we need to “set aside” *all* of our evaluative beliefs. To evaluate Joyce’s EDA, we need to “set aside” *all* of our moral beliefs, including those involving what morality is about. And since we cannot set aside the beliefs in question in either case, it becomes impossible to evaluate their arguments. Thus, Vavova’s claim is about the rationality of disagreement between interlocutors and the limits of rational belief revision, rather than about the logical structure of arguments.

First, it is surely possible to consider an ambitious EDA as involving a contradiction. For example, Street believes that evolutionary influence on evaluative beliefs makes them epistemically suspect. Since her target is all evaluative beliefs, her empirical premise implies that there is an evolutionary influence on our beliefs about epistemic reasons too. So, according to Street’s reasoning, our beliefs about epistemic reasons *are* epistemically suspect. But she also takes epistemic reasons for granted when she gives the evidence of error. This means her argument also implies that our beliefs about epistemic reasons *are not* epistemically suspect. This looks like a contradiction to me.

Second, Vavova’s argument does not really do much if it is purely about “setting beliefs aside.” Even if we accept that Street is making an illegitimate epistemic move, this does not take away from the intuitive strength of her argument. This is because we take many of our beliefs about evidence and belief revision for granted, just as we take many of our perceptual or mathematical beliefs for granted, unless we are all dedicated

skeptics. These beliefs constitute an *independent* ground on which we evaluate other beliefs. They constitute an independent ground because fields such as mathematics, science, and epistemology are independent fields with internal standards of justification and criticism. The contents of these fields go beyond evolutionary influence. Therefore, regardless of whether Street can set aside our beliefs about epistemic reasons or not, we find evaluative realism implausible if we also accept that there is an immense evolutionary influence on our evaluative beliefs.

This shows the real problem of Street's argument (and, for that matter, all ambitious EDAs). Street, like the rest of us, takes our beliefs about epistemic reasons for granted when she gives the evidence of error. Taking such beliefs for granted, in and of itself, is not deeply problematic since certain truths such as "3+4=7" or the truth of certain basic epistemic principles are not affected by the biological evolution of living organisms. However, this move contradicts Street's *empirical* premise because Street, by taking our beliefs about epistemic reasons for granted, implicitly accepts that the content of some of our evaluative beliefs is not directly determined by evolutionary forces. But then the same could go for some of our other evaluative beliefs that she thinks are distorted, including many of our moral beliefs. Morality, as the realist asserts, could be an independent field just like science, mathematics, and epistemology (or philosophy in general). Rejecting this possibility means rejecting evaluative realism from the outset: Street's conclusion is that evaluative realism is implausible, but her empirical premise simply *rejects* evaluative realism. This begs the question against the realist.

2.1.4 Conclusion

I have rejected Vavova's claim that Street's and Joyce's EDAs target too much and become self-refuting. This is because EDAs are essentially *inductive* arguments with a *probabilistic* conclusion, which allows *some* of our beliefs to be true. The "Inverse Rule of Debunking" is true, but not for the reasons Vavova provides. In other words, the level of an EDA's ambition affects the argument's strength but not because more ambition causes an internal contradiction. Rather, the more set of beliefs an EDA calls into question the harder it becomes to provide a complete evolutionary origins story. The empirical premise is the chink in an EDA's armor.

2.2 Greene and the Weakness of the Empirical Premise

An out-of-control trolley is speeding toward five workers who are not able to move out of the way in time. You see a switch connected to the tracks that, if pulled, will divert the trolley onto a sidetrack and kill a single worker who is standing there. Would you pull the switch to sacrifice one person to save five? Or, would you push a large man over a footbridge into the path of the trolley to stop the trolley and save those workers? Now, suppose that you are on a very crowded bus and are struggling to get to the door at your stop. All of a sudden, you notice a man trying to get on the bus, and you suspect that there are explosives strapped around his chest. You realize that the only way to stop the seeming bomber is to push a large man beside you in his direction, so that both will end up on the empty sidewalk and the people on the bus will be spared if the bomb goes off. Would you sacrifice the large man to save the people on the bus?

People often react differently to the above moral dilemmas, even though they all involve sacrificing an innocent person to save more innocent people. For example, most people intuitively accept *pulling a switch* and deny *pushing* as a morally permissible way of sacrificing an innocent person to save more innocent people, when they are presented with the trolley and footbridge dilemmas (Greene et al. 2001). Furthermore, even though both the footbridge and Bus dilemmas involve *pushing* an innocent person to his death to save more innocent people, most people react negatively to the sacrificial action in the footbridge dilemma, yet they find pushing the large man in the Bus dilemma morally acceptable (Railton 2014, 854–5). What would explain these different reactions?

Joshua Greene and his colleagues (2001, 2009) assert that most people find pushing the large man off the footbridge morally unacceptable because their evolutionary-based negative emotional reactions to actions involving “up close and personal” harm *distort* their judgments. Moreover, Greene holds that our negative reactions to actions involving personal harm result in deontological judgments (Greene 2008, 39). Since our deontological judgments have this epistemically suspect, evolutionary origin, they are *normatively inferior* to our consequentialist judgments, which arise from more rational thought processes such as cost-benefit analysis (ibid., 36).⁸

Greene’s explanation for our different reactions to morally indistinguishable dilemmas is not satisfactory because it does not apply to the asymmetry in intuitive moral acceptability in Bus and footbridge dilemmas. It could be true that people often give an evolutionary-based, negative emotional reaction to the sacrificial action in the footbridge

⁸ A deontological judgment roughly takes the form “*X* is wrong *per se*, independent of any consequences it may lead to,” whereas a consequentialist judgment takes the form “*X* is right because its outcome promotes aggregate happiness/utility.”

case when they compare it to the one in the trolley case. Nevertheless, as Railton's experiment reveals, people often find it *morally permissible* to *push* the large man to his death in the Bus dilemma, despite the obvious presence of personal harm.⁹ Thus, the effect of personal harm on our moral judgments cannot explain a general positive attitude towards sacrificing the large man in the Bus case. *What* would explain that positive attitude? And what normative and metaethical implications would *that* explanation have? This section focuses on these questions.

In answering these questions, my general aim is to show why the empirical premise of an EDA is not strong as the debunker believes it is. As I mentioned in 2.1, all EDAs, regardless of their ambition, share a common structure. Thus, it will be sufficient to pick one type of EDA and reveal the weakness of its empirical premise. My discussion is centered on Joshua Greene's EDA, which attempts to undermine the justification of our deontological beliefs. However, my argument works not only against Greene's EDA but also against Street's and Joyce's EDAs.

In 2.2.1, I lay out Greene's main argument and the results of his 2001 study, and in 2.2.2, I present Greene's revised position, which is the result of his 2009 study. In 2.2.3, I propose that different attitudes towards the Bus and footbridge dilemmas are due to the fact that the Bus dilemma is set in a relatable and representative context unlike the footbridge dilemma. I argue that the meager context of the footbridge case (and trolley-type cases in general) prevents many of us from evaluating the situation properly and equalizing the worth of innocent lives, whereas the more relatable and representative context of the Bus case makes it easier for us to make the moral distinction between

⁹ More on Railton's experiment below.

‘harming an innocent person’ and ‘directing a public threat to a lesser harm.’ This enables us to offer principled reasons for why the personal harm in the Bus case is morally permissible, thus justifying our positive (and consequentialist) reaction. In some other representative contexts, though, we often employ certain moral distinctions, such as between ‘harming an innocent person for pleasure’ and ‘increasing the overall happiness,’ to justify our deontological responses.

To account for the way we generally justify our moral claims, I distinguish, in 2.2.4 and 2.2.5, between different types of moral intuitions and claim that our theoretical moral intuitions, as opposed to concrete and mid-level ones, are independent of direct evolutionary influence because they are the product of autonomous (gene-independent) moral reasoning. Since both consequentialist *and* deontological theoretical intuitions are immune to nonmoral biases, it would be wrong to make a substantive normative distinction between the two.

Finally, in 2.2.6, I describe how the exercise of moral reasoning and the process of cultural evolution could generate gene-independent consequentialist *and* deontological moral intuitions that allow us to grasp objective moral facts and distinctions.

This section has both specific conclusions and a more general conclusion. While specific conclusions pertain to Greene’s argument, the general conclusion is more relevant to my position in this chapter, namely that EDAs, and evolutionary considerations in general, are not strong enough to undermine moral realism or Kantian constitutivism. My three specific conclusions are, (1) the effect of personal harm on our moral judgments fails to provide an adequate explanation of the different reactions to some morally indistinguishable dilemmas; (2) the theory I support offers a better

explanation of our consequentialist and deontological responses to particular cases than Greene's theory; and (3) Greene is not justified in his claim that deontology is normatively inferior to consequentialism. But more importantly, my general conclusion is that evolutionary explanations are insufficient to debunk some of our moral beliefs and intuitions. Thus, EDAs fail to achieve their mission.

2.2.1 Greene's Argument

Greene claims that consequentialism is normatively superior to deontology. To bolster his claim, he aims specifically to undermine the justification of our deontological beliefs by placing a focus on their evolutionary origins. That is, on his account, deontological judgments are typically fueled not by the rational pull of objective moral facts but by morally irrelevant emotional responses which evolved as a means of increasing biological fitness. These emotional responses, according to Greene, are more likely to occur when one finds oneself in a situation where there is a *personal* harm rather than an *impersonal* one.¹⁰ Greene's important claim is that whether the harm is "up close and personal" or impersonal is morally irrelevant. This means, according to Greene, only the harm itself must have moral importance. Thus, strong negative responses to cases involving personal harm cannot be the result of the rational force of a deontological principle. Rather, they must be rooted in our evolutionary past: "These responses evolved as a means of regulating the behavior of creatures who are capable of intentionally harming one

¹⁰ An action involves *personal* harm if it is "(i) likely to cause serious bodily harm, (ii) to a particular person, (iii) in such a way that the harm does not result from the deflection of an existing threat onto a different party" (Greene and Haidt 2002, 519).

another, but whose survival depends on cooperation and individual restraint” (Greene 2008, 43).

Greene bases his claim on the results of a study he and his colleagues conducted regarding responses to various moral scenarios, including the famous trolley dilemma.¹¹ Greene et al.’s (2001) study shows that whether the harm is “up close and personal” affects people’s moral judgments. In the study, subjects are presented with both personal and impersonal moral dilemmas and their brain activity is measured using fMRI when they respond. The results are interesting: (a) when subjects are presented with a personal dilemma like the footbridge case, higher activity in brain regions connected with emotional response is observed compared to their brain activity when presented with an impersonal dilemma like the trolley case; (b) people who, contrary to the majority opinion, decided to push the large man in the footbridge case took longer to finalize their decisions than people confronted with the trolley case (ibid., 2017). Greene infers two conclusions from these findings: (1) people’s emotional responses to morally indistinguishable dilemmas differ according to the way harm being inflicted; (2) the longer response time shows that the cognitive system in the subjects’ brain *overrides* the emotional system and results in the decision to push the large man.¹² Therefore, Greene

¹¹ The trolley dilemma was introduced by Philippa Foot (1967).

¹² According to Greene, two psychological systems compete to produce a moral judgment: emotional and cognitive systems. The emotional system mostly involves quick, automatic, but not necessarily conscious responses. We mostly resort to our emotions to survive, for they are “very reliable, quick, and efficient responses to recurring situations” (Greene 2008, 60). By contrast, the cognitive system involves processes such as “reasoning, planning, manipulating information in working memory, controlling impulses, and higher executive functions” (ibid., 40). These processes do not trigger particular behavioral responses like emotions do. Rather, they are neutral since they take many factors into account and any change in these factors could lead to a different course of action.

and his colleagues think that “the crucial difference between the trolley dilemma and the footbridge dilemma lies in the latter’s tendency to engage people’s emotions in a way that the former does not” (ibid., 2106).

Greene thinks that consequentialism is normatively superior to deontology because he associates deontological judgments with the emotional system and consequentialist judgments with the cognitive system, and more importantly, he claims that the emotional system that produces deontological judgments are *distorted* by morally irrelevant factors such as the proximity of harm. Deontologists usually concede the importance of producing best consequences, as they do in the trolley case. Nevertheless, they sometimes choose to place constraints on consequentialism on account of rights or intrinsic human dignity – constraints which apply in the footbridge case. Greene et al.’s study suggests that deontologists offer these constraints not due to the rational force of a deontological principle but due to their strong negative reaction to instances involving personal harm.

Greene’s argument can be summarized as follows:

(P1) The emotional system that induces deontological judgments is affected by features that render a moral dilemma personal.

(P2) The features that render a moral dilemma personal are morally irrelevant.

(C1) Therefore, the emotional system that induces deontological judgments is affected morally irrelevant features. (P1, P2; empirical conclusion)

(P3) Moral judgments that are affected by morally irrelevant features do not reliably track moral facts.

(P4) Deontological judgments do not reliably track moral facts. (C1, P3)

(P5) Moral judgments have a genuine rational or normative power only if they reliably track moral facts.

(C2) Therefore, deontological judgments, in contrast to consequentialist judgments, lack a genuine rational or normative power. (P4, P5; philosophical conclusion)

2.2.2 Personal Force

There is a quick objection to Greene's first premise. Suppose that we change the conditions in the footbridge case: as in the footbridge case, a large man is standing on a footbridge, but you are standing near a remote switch. If you pull the switch, you will knock the victim off the bridge by activating a trap door. Let's call this case *Remote Footbridge*. In Remote Footbridge, we are still using someone as mere means but now the harm appears to be *impersonal*. If this variation does not change our reaction, then Greene is mistaken: the nature of our reaction to footbridge cases does not depend on whether the harm is personal. Rather, we have strong emotional responses to such cases because we have *good reasons* to believe that killing an innocent person is wrong.

Greene has a good answer to this objection. In their 2009 study, Greene et al. introduce variations of the footbridge case including Remote Footbridge. The aim of the study is to improve on the 2001 study by making pairwise comparisons between variations of the footbridge case and trying to find out which morally irrelevant factor we respond to. There are three morally irrelevant factors in the standard footbridge dilemma: (i) *spatial proximity*, (ii) *physical contact*, (iii) *personal force*. Personal force differs from physical contact in that personal force occurs "when the force that *directly* impacts the other is generated by the agent's muscles, as when one pushes another with one's hands or with a rigid object" (Greene et al. 2009, 365). Greene and his colleagues introduce three variations in addition to the standard case:

Remote Footbridge: Everything is the same except that you knock the victim off the bridge using a trap door and a remote switch. This action involves *none* of the morally irrelevant factors.

Footbridge Pole: Everything is the same except that you use a pole instead of your hands to push the victim. This action involves *spatial proximity* and *personal force*.

Footbridge Switch: Identical to the Remote Footbridge except that you and the switch are next to the victim. This action only involves *spatial proximity*.

Comparing Remote Footbridge to the Footbridge Switch isolates *spatial proximity*.

Comparing Standard Footbridge to Footbridge Pole isolates *physical contact*. And comparing Footbridge Switch to Footbridge Pole isolates *personal force*.

The results reveal that there is *no* significant effect of *spatial proximity* (Remote Footbridge vs. Footbridge Switch) and *physical contact* (Standard Footbridge vs. Footbridge Pole) on our responses, but there is a significant effect of *personal force* (Footbridge Switch vs. Footbridge Pole).¹³ That is, we are inclined to find harmful actions involving personal force morally wrong. Greene thereby shows that we generally do not react negatively to actions that do not involve personal force such as the actions in the Remote Footbridge and Footbridge Switch. The objection mentioned above, thus, collapses.

2.2.3 Explaining Different Reactions

It could be true that we generally do not react negatively to actions that *do not* involve personal force. However, *do most people react negatively to actions involving personal force on deontological grounds?* If we consistently found most people reacting negatively to actions involving personal force and consequently making deontological judgments, then Greene's first premise would seem plausible. But if most people *did not* react negatively to such actions and made consequentialist judgments instead, then it would be wrong, or at least hasty, to assume a strong correlation between deontological

¹³ Cf. Greene et al. 2009, 367.

judgments and actions involving personal force. A negative answer to the above question is suggested by another moral dilemma put forward by Peter Railton:

The Bus Dilemma: You live in a city where terrorists have in recent months been suicide-bombing buses and trains. The terrorists strap explosives to themselves under their clothing, and, at busy times of the day, spot a crowded bus or train and rush aboard, triggering the bomb instantly to avoid being stopped. You are on a very crowded bus at 5:10 pm, and are struggling to get to the door at your stop. The doors are starting to close and you won't be able to get off unless you jostle the slow-moving obese gentleman trying to exit at the same time. Suddenly you notice a man rushing up to the bus and forcing his foot into the doorway, wedging it between the fat man and the door frame. He is reaching with one hand under his coat and a gap between the buttons reveals to you what look like explosives strapped around his chest. You can't reach this man, but if you push the corpulent gentleman beside you hard in his direction right now, he will fall directly on top of the seeming bomber and both will end up on the empty sidewalk, while you fall backwards into the bus as the doors snap shut. So, if you push hard, and this man is not a bomber, then the bus will leave behind two very annoyed men on the sidewalk, and you will be left on the bus, covered with embarrassment. But if he is a bomber, the bus will be spared, and you with it, but the fat man killed as the bomber explodes underneath him. On the other hand, if you simply squeeze off the bus alongside the corpulent gentleman and do nothing more, and the other man is a bomber, then many people on the bus will be killed while you and the corpulent gentleman are safe on the sidewalk. But if this man is not a bomber, then no one on the bus will be hurt and you simply will have jostled a corpulent gentleman while exiting a bus, and you can apologize to him on the sidewalk. Whatever happens, you will not be killed if there is a bomb and it goes off—you will either be on the bus when it explodes on the sidewalk, or on the sidewalk when it explodes on the bus. Should you (a) shove the corpulent gentleman hard right now, or (b) squeeze off the bus, jostling the corpulent gentleman but doing nothing else? (Railton 2014, 854–5)

Railton uses the Standard Footbridge, Footbridge Switch, and Bus dilemmas to conduct an uncontrolled and unscientific experiment. According to the results, 71% of the participants find pushing the large man off the bridge morally wrong (Footbridge), and 72% of them find pulling the switch morally acceptable (Footbridge Switch). More importantly, most of the participants (67%) also find pushing the corpulent gentleman morally acceptable (Bus).

Why do the same people judge one instance of personal force to be morally *impermissible* on *deontological* grounds and another to be morally *permissible* on *consequentialist* grounds, even though *both* instances involve sacrificing an innocent person to save more innocent people? I suggest that the reason for our differential responses is that the Bus dilemma is set in a relatable context, as opposed to the footbridge dilemma. We can imagine the Bus case more vividly than the footbridge case: we are among the passengers who might get killed, there is a common awareness of terrorist attacks, we can feel a sense of social solidarity and collective self-defense.¹⁴ By contrast, no social background is given in the footbridge case and thus the situation strikes us as improbable. Since the life-or-death pressure in the Bus dilemma cannot easily be simulated in our brain when we are confronted with the footbridge dilemma, people understandably have more motivation to push the large man in the Bus case. It is also easier in Bus case to imagine ourselves being able to cope with the feelings such as guilt, regret, and shame after pushing the victim. And it is harder in Bus case not to be haunted by guilt, regret, and shame after failing to push him. Therefore, it seems that we are more prone to distortions arising from our evolutionary past in our responses to trolley-type cases involving personal force. It seems that the meager context of the footbridge case (and trolley-type cases in general) prevents many of us from evaluating the situation properly and equalizing the worth of innocent lives. Since we can evaluate the situation properly in the Bus case, it becomes easier to empathize with those we might save and to equalize the worth of innocent lives. Thus, we often react positively to actions in which we use personal force to save more innocent lives.

¹⁴ Cf. Railton 2014, 857.

The representative context of the Bus dilemma enables most of us to make the moral distinction between ‘harming an innocent person’ and ‘directing a public threat to a lesser harm.’ We can thereby provide *reasons* to justify our *consequentialist* response (or support) to the use of personal force in that case. That does not entail, however, that we are only able to give consequentialist justifications in representative contexts. It is perfectly possible for us to make moral distinctions to justify our *deontological* responses to some other relatable and representative situations.

Consider the following case, which we may call ‘The Sadistic Case.’ A homeless person with no family or friends arrives in a new city with the hope of a better life. He falls asleep shortly after he finds a comfortable spot to relax. He wakes up to find his hands and feet bound, surrounded by a group of masked sadists who start hitting him with sticks. His screaming and begging for mercy only intensify the abuse and increase the pleasure of the delighted torturers. One of the torturers is live streaming the whole event on the Dark Web so that thousands of sadists from all around to world can watch it and share the torturers’ pleasure without being caught. They successfully dispose of the body after he dies, and no one finds out about the incident over the following years. Everyone involved remains silent about what happened and are never discovered. Killing an innocent person in the Sadistic Case strikes many of us as wrong, unlike in the Bus case. And when we are asked to justify our reaction, we can point to the moral distinction between ‘harming an innocent person for pleasure’ and ‘increasing the overall happiness.’ In other words, we can justifiably claim that killing the homeless person is morally wrong despite the increase in the overall happiness.

2.2.4 Types of Moral Intuitions and Autonomous Moral Reasoning

2.2.4.1 Types of Moral Intuitions

As I argued above, Greene's account of the effect of personal force on our moral judgments does not provide a satisfactory explanation of our consequentialist and deontological responses to moral dilemmas that are set in relatable and representative contexts and of the legitimate ways we can justify those responses. Michael Huemer's (2008, 383–6) categorization of moral intuitions¹⁵ shows how we generally justify our consequentialist and deontological reactions to relatable and representative cases. It also shows that Greene is not entitled to associate selective pressures with our *theoretical* moral intuitions, be they consequentialist or deontological.¹⁶ On Huemer's account, there are three types of moral intuitions:

- (1) *Concrete intuitions*: Intuitions about particular cases such as trolley and footbridge dilemmas.
- (2) *Theoretical (abstract) intuitions*:¹⁷ Intuitions about general moral principles such as the consequentialist intuition that the right action is the one whose consequence promotes the overall happiness, or the deontological intuition that it is wrong to use people as mere means.

¹⁵ By intuition, I mean “immediate intellectual grasp or appearance prior to reasoning.” There are roughly two ways of conceiving intuitions. We can see them as a species of belief with a distinct kind of justification. For example, once we understand the proposition “20 is less than 100,” we are automatically justified in believing it. This does not require reasoning or inference but merely understanding the proposition. Or we can see intuitions as cognitive states different from beliefs. Considered in this way, an intuition takes the form “It seems to me that *X*,” which is different from having a belief. Our beliefs may differ from our intuitions: a counterexample to an intuition or an evidence against it could make us disbelieve a certain intuition (cf. Huemer 2008, 370).

¹⁶ Huemer (2008) uses his categorization of moral intuitions to support his ‘revisionary intuitionism,’ while I use it against Greene’s argument and reach different conclusions than Huemer. I also provide an account of how we acquire theoretical intuitions in the sixth section, which, I believe, is necessary to be able to classify them as ‘intuitions.’ I am not sure if Huemer would accept that account or my conclusions.

¹⁷ Huemer calls them ‘abstract intuitions,’ whereas I prefer to call them ‘theoretical intuitions.’

(3) *Mid-level intuitions*: Intuitions about principles that have in-between level of generality such as the principle that one ought not to kill sentient beings for entertainment, or that one ought not to lie even if it is the only way to save one's life.

Huemer claims that concrete and mid-level intuitions are more likely to be responsive to evolutionary biases, whereas abstract or theoretical intuitions are probably not directly responsive to them. We can give two reasons for this conclusion: (1) concrete and mid-level intuitions are related to strong emotional reactions, and (2) evolutionary mechanisms select for certain types of behavior to promote biological fitness.

Consider a concrete intuition such as that "It is wrong that David had sex with his full sister," or a mid-level intuition that "It is not permissible to kill sentient beings for entertainment." These intuitions are more likely to arouse strong emotional reactions associated with our evolutionary past than the theoretical intuition, say, "*S* is morally obligated to do *x*, only if *S* is capable of doing *x*." Furthermore, evolution might have endowed us with intuitions that favor certain types of behavior that are more likely to increase our chances of survival and reproductive success. These could include, for example, intuitions about incest, infanticide, murder, and promiscuity among many others. However, it is less likely that it has endowed us with intuitions that favor certain types of abstract principles. Admittedly, there are studies that indicate our ancestors' involvement in moral reflection to some extent;¹⁸ nevertheless, it is less than obvious that there is a meaningful relation between theoretical intuitions and reproductive success. It is true that some of our shared evaluative dispositions may have arisen from properties conducive to survival. For instance, properties such as fractal patterns and symmetry

¹⁸ Cf. Flack and de Waal (2000), and Burkart et al. (2018).

could be the evolutionary foundations for our appreciation of beauty.¹⁹ Theoretical intuitions, by contrast, do not seem to be very helpful in terms of survival. Abstract moral principles (e.g., one should always maximize utility, or one should never use people as mere means) are *too general* to qualify as useful guides to survive and reproduce. It is likely that we reach abstract moral principles through employment of autonomous reflection on our evolved reactive attitudes to particular cases. And over time, we convert these principles into intuitions and internalize them, as I describe it in the sixth section.

One immediate objection to the idea that theoretical intuitions are not useful in terms of survival and reproduction is as follows. Imagine a community that is not getting along, constantly fighting with each other and lowering overall utility. The utilitarian local leader says, “We have to get along and make as many people happy as we can. Let’s maximize overall happiness!” Similarly, the proto-Kantian group leader could say, “We need some rules that everyone can follow; otherwise, we won’t agree on anything. If each person just does what they desire, we will be in conflict. So, let’s at least reject principles that not everyone could follow, then we will get along.” One could argue that these folksy versions of utilitarianism and Kantianism would help with survival and reproduction. Admittedly, following rules about particular actions or particular kinds of actions could be helpful. For example, rules such as “Don’t lie to Eliana about last night,”

¹⁹ Fractal patterns occur in clouds and waves. Trees grow symmetrically and antlers of a deer are symmetrical. These properties helped our hominin ancestors evaluate their environment more easily and react quickly to danger. For example, they learned to avoid deformed plants because they might not be safe to eat (cf. Freeman et al. 1993). They preferred mating partners with a symmetrical face because a symmetrical face is an indication of health and fertility (cf. Chatterjee 2014, ch.3; Jasienska et al. 2006). In other words, these properties, among many others, have become signals of safety and nutrition, triggered nice feelings in us, and consequently lead to our appreciation of beauty. This could be a possible explanation of how our concept of beauty might be related to evolutionary forces.

“Never make any sexual advance to those not consenting,” “Don’t kill/torture/lie/rape/steal etc.” could promote survival and reproduction if people follow them most of the time. Nevertheless, people often disagree on particular applications of abstract moral principles. Some would think following the rule *A* is the correct application of the moral principle *P* in circumstances *C*, whereas some would think *A* goes against the nature of *P* in *C* and offer an utterly different rule instead. It is less than obvious which specific rules or actions would maximize overall happiness, or which principles could be followed by everyone. Theoretical intuitions alone are not evolutionary advantageous but some of the applications of them to particular (kinds of) situations could be.²⁰

Is the distinction between different types of moral intuitions too vague? We arguably need specific criteria that an intuition must meet for us to classify it reliably as theoretical, mid-level or concrete. I propose *three* criteria which an intuition must meet to qualify as a theoretical intuition: (1) it does not arise from specific set of environmental circumstances of time and place (e.g., trolley cases), (2) it does not require acting in a specific way (e.g., pulling a switch), and (3) it does not involve a particular person (e.g., large man). Let’s look at some examples:

(D₁) It is unfair when some people are worse off than others owing to differences in their unchosen circumstances.²¹

²⁰ Likewise, scientific theories alone are not evolutionary advantageous but certain applications of them to particular (kinds of) situations could be. Scientific and technological advances could potentially wipe out all the humans on Earth, or they could extend the average human lifespan, depending on how we apply them to particular situations.

²¹ This is the principle of luck egalitarianism, which is against the view that distribution matters only when it promotes the overall well-being. This echoes the intuition that fairness matters irrespective of the consequences. Without a doubt, there could be sophisticated consequentialist theories that address our intuitions about fairness or justice. However, this would simply mean that those sophisticated theories are partly shaped by *deontological* theoretical intuitions.

(D₂) It is our duty to fulfill a promise, regardless of the goodness of the consequences.

(D₃) It is morally wrong to prosecute and punish those known to be innocent.²²

(C₁) An action is morally good if it increases the overall level of pleasure in the world.

(C₂) An action is morally bad if it increases the overall level of pain in the world.

While D₁, D₂, and D₃ are deontological theoretical intuitions, C₁ and C₂ are consequentialist theoretical intuitions. These intuitions do not specify any circumstances, any particular action, or any particular individual. Thus, it seems that they are *not* generated directly by our emotional responses to particular cases, but rather by autonomous moral reasoning.

2.2.4.2. *Autonomous Moral Reasoning*

What makes moral reasoning *autonomous* or *gene-independent*? One way to interpret the evolutionary influence on our capacity for morality and the content of our moral judgments is to claim that moral thought and behavior are the direct result of evolutionary mechanisms and have the sole function of adapting humans to their changing surroundings.²³ This strictly behavioristic and eliminativist view sees morality as a “certain type of behavioral pattern or habit, accompanied by some emotional responses” rather than as “a theoretical inquiry that can be approached by rational methods, and that

²² This is Shafer-Landau’s (2003, 248) example.

²³ For example, Michael Ruse’s Darwinism says that “substantive [or objective] morality is a kind of illusion, put in place by our genes, in order to make us good social cooperators” (Ruse 2010, 309).

has internal standards of justification and criticism” (Nagel 1979, 142). We can call the latter the *theoretical* conception of morality, as opposed to the *behavioristic* conception.

The idea behind the behavioristic conception of morality is that evolutionary processes cause us to adopt certain moral beliefs that are consistent with the purpose of natural selection, namely, to maintain and promote survival and reproductive success. Adopting these beliefs enables cooperation between beings with selfish desires, and groups that cooperate have more reproductive success than groups that lack cooperation.²⁴

Evolutionary advantages of cooperation explain the similarity between humans and nonhuman animals in terms of altruistic behavior. Many nonhuman animals exhibit altruistic behavior such as parental care, food sharing, and coalition formation. There is a high degree of division of labor in colonies of social insects such as bees and ants, and when necessary, self-sacrifices are made by individual bees or ants to defend the colony. Hamilton (1964) explains altruistic behavior of social insects by appealing to *kin selection*, which supports the idea that such behavior is genetically determined.²⁵ If altruistic behavior of nonhuman animals is genetically determined, we should be able to make the same claim about human altruistic behavior, since humans are animals too. And if natural selection shapes the capacity for morality and the content of our moral beliefs in this *direct* way, then it seems safe to claim that “the time has come for ethics to be removed temporarily from the hands of the philosophers and biologicized” (Wilson 1975, 562).

²⁴ Cf. Darwin 1981, 166.

²⁵ For example, worker bees have evolved to act selflessly in order to serve the queen’s reproductive success because (i) each of them shares some of their genes with the queen, and (ii) they are sterile.

Not all evolutionary biologists think that the capacity for morality and the content of our moral beliefs are both biological adaptations. Prinz (2008) and Ayala (2016) argue that the capacity for making moral judgments and normative guidance is not directly shaped by natural selection, but rather it is a byproduct of non-moral human intellect. Ayala also claims that most of the content of morality is not biologically determined but is a result of cultural evolution (Ayala 2016, 258–60). According to Ayala, human moral behavior is not causally related to prosocial behavior of nonhuman animals. Rather, moral behavior amounts to *rational* behavior in our case because our capacity to make moral judgments is the product of our intellectual powers such as *self-awareness* and *abstract thinking* (Ayala 2016, 250).

Ayala talks about three necessary conditions for moral behavior: (1) “the ability to anticipate the consequences of one’s own actions,” (2) “the ability to make moral judgments,” and (3) “the ability to choose between alternative courses of action” (Ayala 2016, 246). Having these abilities enables us to think about moral propositions, evaluate them, and act accordingly. However, only beings with an advanced intelligence can have the capacity for morality.²⁶ Therefore, animals that lack such an intelligence do not really exhibit proper moral behavior, but their behaviors merely express their genes. On the other hand, we, unlike nonhuman animals, make moral judgments as a result of our advanced intellectual capacities and we carry out an evaluation when we think about those judgments. That is, our advanced intelligence allows us to assess various moral

²⁶ In Ayala’s view, there is an evolutionary threshold that should be crossed in order to meet the necessary conditions for moral behavior. Although it may be very difficult to determine when our ancestors actually reached the threshold, we can be sure that no proper moral behavior is possible unless the threshold is crossed. We can liken this evolutionary threshold to the boiling point of water (cf. Ayala 2016, 250).

beliefs and decide whether they are beneficial to society and its members (Ayala 2016, 259).

The theoretical conception of morality does not deny the fact that we and some species of nonhuman animals share certain basic psychological dispositions and emotional responses that have been shaped by evolutionary forces. Its distinctive claim is that we also possess a superior intelligence that allows us to evaluate, systematize, and occasionally say no to these pre-reflective dispositions and emotional responses. If Ayala is right in his claim that we acquired the capacity for morality through our advanced intelligence, it may be true that we exercise autonomous reasoning in reaching some of our moral judgments. That is, we do not always automatically act on our adaptive dispositions or emotions, but we also think about them, evaluate them, try to justify them by appealing to reasons, decide between them, and guide our actions in line with the judgments reached through rational reflection. Moral reasoning is autonomous, then, in the sense that moral thought is not significantly determined by specific, evolutionary-based, psychological dispositions. For example, it may turn out that dispositions that promote slavery, racism, and condemnation of homosexuality are adaptive. But after exercising moral reasoning, revising our judgments about these cases, and passing down our revised judgments to subsequent generations, many people have come to suspect that beliefs that support slavery, racism, and discrimination against LGBTIQ+ people are unjustified, and accordingly they have been trying to act against these adaptive dispositions. The fact that we have the capacity to criticize and revise our adaptive responses supports the idea that human morality is not just a totality of behavioral patterns but also a theoretical inquiry.

Denying the influence of autonomous reasoning on our moral beliefs leads to a hyperselectionist view that all our intellectual activities ultimately aim at evolutionary ends. According to such a view, capacities such as abstract thinking, intentionality, and self-awareness are essentially not different from patterns of behavior in that they are all affected by basic psychological dispositions and geared towards cooperation. It follows from this strictly behavioristic view of cognition that no thinking occurs independent of genes. It is highly problematic to assert that each (cognitive) trait contributes to biological fitness. This is because of the explanatory gap between genotype and phenotype. There is no direct correspondence between genotype and phenotype: (1) There are traits that are irrelevant to survival and reproductive success such as vestigial organs;²⁷ (2) There are inheritable diseases such as schizophrenia and Alzheimer's disease, which reduce biological fitness; (3) There are pleiotropic traits, which show us that adaptive changes could generate traits that are not themselves adaptive;²⁸ (4) There are *exaptations*, shifts in the function of a trait in the course of evolution, such as plumage of birds.²⁹

The fact that pre-reflective, adaptive dispositions cannot determine the content of certain mental activities points to the human capacity for autonomous (gene-independent) reasoning. For example, it would be a mistake to explain our mathematical or scientific judgments only by appeal to their psychological origins. The *reasons* behind those judgments should also be taken into account. We can justify our mathematical, logical, or

²⁷ On traits that are not affected by natural selection, see Flanagan 1991 and Collier and Stingl 1993.

²⁸ Pleiotropic genes are the genes that influence multiple unrelated traits. Pleiotropic traits are unrelated traits that are affected by the same gene (cf. Paaby and Rockman 2013).

²⁹ Cf. Gould and Vrba 1982, 5–7.

scientific beliefs only when they correspond to mathematical, logical, or scientific *facts*, many of which are independent of how evolutionary mechanisms work. For instance, truths of the propositions “ $1+1=2$ ” and “Water boils at 100°C ,” and the principle of non-contradiction are independent of how organisms evolve and what kind of basic dispositions they possess. Although it may be true that our general disposition to become involved in mathematics, logic, and science is an adaptation, it is less likely that evolutionary influence is so pervasive as to determine the content of these activities. This indicates a shift in the function of our intellectual capacities: they do not only facilitate survival and reproduction, but they also help us track the truth of certain objective facts that go beyond the workings of evolutionary mechanisms.

Moreover, even though moral codes must be consistent with our biological nature,³⁰ “moral norms are independent of [biologically conditioned human] behaviors in the sense that some norms may not favor and may hinder [survival and reproductive success]” (Ayala 2016, 245). For example, religion and patriotism facilitate cooperation but they also give rise to beliefs about racism and genocide, which may act against biological fitness. Obedience is considered as being conducive to survival³¹ but it can turn into mass killings in the wrong hands. The rapid decline in birth rates in some areas of Italy in the nineteenth century is another example of human values going beyond evolutionary aims.³² All these examples point to the fact that *some* of the content of our moral beliefs

³⁰ Otherwise, our species would cease to exist. For example, if killing were the norm, there would ultimately be no human being to think about morality.

³¹ See Darwin 1981, 166.

³² See Cavalli-Sforza and Feldman 1981.

are *exaptations* rather than *adaptations*. Thus, we have good reasons to claim that human moral behavior differs significantly from altruistic behavior of nonhuman animals due to the distinct human (or rational) capacity for *reasoning* and *evaluation* and the effect of cultural transmission.³³

One might still object to the theoretical conception of morality by pointing to the differences between morality and subjects like mathematics and science. Such an objection, *prima facie*, sounds reasonable because while the main objective of mathematics and science is to have an accurate picture of the world, the main objective of morality is to guide our actions. That is, morality has the *practical* aspect mathematics and science lack. However, this difference does not necessarily render the employment of autonomous moral reasoning improbable. In the case of morality, we do not only try to justify our *beliefs* like we do in mathematics and science but also our *desires* and *dispositions* to engage in certain acts.³⁴ As long as we can subject our evolved psychological desires and dispositions to scrutiny and decide whether we are justified in acting in accordance with these desires and dispositions, it is perfectly possible that we employ autonomous reasoning in making and revising moral judgments as well.

Greene does not deny that moral reasoning is autonomous. Rather, on his account, autonomous moral reasoning favors consequentialism because he believes that his studies indicate that deontological judgments, unlike consequentialist ones, are shaped by our evolutionary-based emotional reactions to morally irrelevant factors. My point is that the nature of our consequentialist *and* deontological theoretical moral intuitions indicates that

³³ More on the effect of cultural transmission below.

³⁴ Cf. Nagel 1979, 144.

they *both* are the result of *autonomous* application of human intelligence in moral thinking.

Turning back to theoretical moral intuitions, one important objection to the contrast between emotional responses to particular cases and reasoning about abstract theories is the claim that *all* abstract moral theories ultimately rest upon our responses to particular cases; thus, they are subject to evolutionary influence as well. That is, we first have emotional reactions to particular cases. Then we reflect on these reactions and come up with abstract moral theories. Since these abstract theories originate from our evolved reactive attitudes, they also must have been shaped by natural selection. This is the GIGO problem,³⁵ but it applies to both consequentialist and deontological theoretical intuitions. *If* deontological theoretical intuitions are epistemically suspect because they are distorted by our evolved attitudes to particular cases, so are consequentialist theoretical intuitions because they arise from our attitudes to particular cases as well. For instance, it is highly unlikely that we arrived at C₁ and C₂ without reflecting on our evolved attitudes to particular cases involving specific kinds of harm. This is one side of the coin: a possible story about how we reach moral judgments. The other side of the coin involves another equally (and perhaps a more) plausible story: the fact that particular cases arouse emotional reactions does not tell us whether the actions in these cases are *morally right, wrong, or neutral*. To decide that, we employ autonomous moral reasoning and decide whether our reactions are *appropriate*. In cases like condemnation of homosexuality, we have come to suspect that our reactions are not appropriate, whereas in cases like looking

³⁵ The idea is that if our basic psychological dispositions are heavily shaped by natural selection, and if our moral reasoning stems from those basic dispositions, then our moral reasoning must also have been shaped by natural selection. Greene calls it “the GIGO problem: garbage in, garbage out” (2008, 116). Street (2006, 124) and Kahane (2011, 119) make the same point.

after our children we believe that our reactive attitudes are on the right track. What I claim is *not* that the former story is absolutely wrong, and the latter is absolutely correct. Although I favor the latter story, my more modest point in this paragraph is that the truth of *either* story would go against Greene's claim that consequentialism is normatively superior to deontology.

Let's reiterate the empirical premise of EDAs: "*Evolutionary mechanisms have a pervasive influence on the content of our evaluative/moral beliefs.*" At first glance, it seems that the empirical premise simply reflects the findings of the studies that merely *describe* evolutionary origins of moral phenomena. Nevertheless, it does more than merely describing the causal relationship between evolutionary forces and moral beliefs. The empirical premise, in fact, adopts a certain metaethical interpretation of the data provided by evolutionary biology: it echoes Ruse and Wilson's *eliminativist* or *behavioristic* approach to morality because of the adjective 'pervasive.' Admittedly, if the truth of the empirical premise were self-evident, then the argument would get inductively very strong and would have the desired rational pull. However, asserting a pervasive influence on the content of our moral beliefs amounts to a rejection of autonomous moral reasoning, which commits the empirical premise to an outright rejection of objective moral values. In other words, since the eliminativist interpretation is not the only available metaethical option, the empirical premise involves an illegitimate assumption that moral realism (or non-naturalist accounts of moral objectivity) is false.³⁶ And resorting to the GIGO problem does not seem to save the

³⁶ It is an illegitimate assumption because the skeptical conclusion of an EDA states that moral realism is implausible. It would beg the question to support this conclusion with a premise that assumes the falsity of moral realism.

debunker. Street, for example, attempts to undermine the power of gene-independent moral reflection by asserting that “[r]ational reflection must always proceed from some evaluative standpoint” (Street 2006, 124). That is, moral reasoning amounts to “assessing some evaluative judgments in terms of others” (ibid.). The problem with Street’s response is that it is founded upon the very assumption that is at issue: that tools of reasoning are contaminated due to the widespread evolutionary influence on our moral beliefs. Street, Greene, and other debunkers need to provide an independent reason to undermine the human capacity for gene-independent rational judgment.

The only weakness of our theoretical intuitions seems to be that they are vulnerable to exceptions.³⁷ This means one could always come up with a counterexample to abstract moral principles, and once a counterexample is found the intuition at issue loses its initial credibility. This occurs quite often in ethical discussions. Deontological ethics is often criticized for being *too restrictive*. For instance, if lying amounts to using another rational being as mere means, then lying must be forbidden in *all* circumstances, including the one in which telling the truth gives rise to death of your children. Once this counterexample is raised, we develop a more critical attitude towards the theoretical intuition that we should never use people as mere means. Consequentialism, on the other hand, is often criticized for being *too demanding*. This is because consequentialism requires us to give away almost all our wealth to the poor, since it will increase the overall happiness.³⁸ For many, this goes above and beyond what morality requires of us. Such counterexamples pose a challenge to generalizations that are reflected in our

³⁷ Cf. Huemer 2008, 384.

³⁸ See Singer 1972, 231–2.

different consequentialist or deontological theoretical intuitions, making it quite difficult to construct a coherent moral system. They encourage revision not only in our theoretical intuitions but also in our mid-level and concrete intuitions.

2.2.5 Formal Intuitions

I have argued that while concrete and mid-level intuitions are susceptible to nonmoral biases, theoretical intuitions seem to be immune to them. At the same time, theoretical intuitions of both consequentialism and deontology are vulnerable to exceptions. Should we now conclude that our moral intuitions are never trustworthy as a guide to practical reasoning? Is there no way forward? Huemer proposes that our most trustworthy intuitions are *formal intuitions*, which are a subset of theoretical intuitions. The function of formal intuitions is not to make any moral evaluation but to place formal constraints on moral theories. Consider three of the examples Huemer gives (2008, 386):

- (a) If x is better than y and y is better than z , then x is better than z .
- (b) If it is wrong to do x , and it is wrong to do y , then it is wrong to do both x and y .
- (c) If two states of affairs, x and y , are so related that y can be produced by adding something valuable to x , without creating anything bad, lowering the value of anything in x , or removing anything of value from x , then y is better than x .

Formal intuitions, according to Huemer, are products of reflection upon what is required by the *nature* of the ‘better than’ relation, wrongness, moral evaluation, permissibility, and so on. While we arrive at theoretical intuitions first by reflecting on particular cases and then reaching a general conclusion, formal intuitions are generated by reasoning about what is entailed by the *nature* of evaluative predicates such as ‘better

than’ or moral concepts.³⁹ In other words, the reasoning from particular cases to theoretical intuitions is *inductive*, whereas the reasoning from formal intuitions to particular moral facts is *deductive*. We can liken formal intuitions to axioms in geometry: we derive particular moral facts from evaluative principles such as the principle of transitivity of ‘better than.’ Since formal intuitions follow directly from the nature of evaluative/moral concepts, they do not arise from observation of particular cases. Therefore, formal intuitions are less likely to be affected by nonmoral, evolutionary biases that are connected with reactive attitudes to particular cases, and they also do not seem to be vulnerable to exceptions.

When a counterexample is found to a theoretical intuition, the intuition loses its initial credibility, which could even result in a rejection of it. However, when one comes up with a counterexample to a formal intuition, we generally call the case a *paradox* instead of giving up the intuition altogether. For example, take Derek Parfit’s famous ‘repugnant conclusion:’

- A*: 100 people live a very high quality of life.
- B*: 200 people live a slightly lower quality of life than the people in *A*.
- C*: 400 people live a slightly lower quality of life than the people in *B*.
- ...
- Z*: 100.2²⁵ people live in conditions barely worth living (Parfit 1984, 419–30).

Most people have the intuition that “*B* is better than *A*,” “*C* is better than *B*,” and so on. And if the principle of the transitivity of ‘better than’ is to hold for our intuitions, then we must expect to have the intuition that “*Z* is better than *A*.” Despite that, most people

³⁹ It is important to note that formal intuitions are not logical laws but principles of ethics. For example, if *A* punches *B* and *B* punches *C*, it doesn’t follow that *A* punches *C*. The schema “*aRb*, *bRc*, therefore *aRc*” is not a valid inference form in general, since it depends on the *nature* of the relation *R*. Transitivity principles are principles of whatever subject matter relation *R* belongs to, not principles of logic.

have the intuition that “*A* is better than *Z*.” We seem to have three options to solve this paradox: (1) reject the transitivity of the ‘better than’ relation, (2) reject some or all our earlier intuitions (e.g., *B* is better than *A*), or (3) accept the “repugnant” conclusion. Denying the principle of the transitivity of ‘better than’ would block the reasoning leading us to the repugnant conclusion at the cost of giving up one of the least controversial evaluative principles. Temkin (1987; 2012) and Persson (2004) espouse such a radical approach. Temkin (2012, 461–5), for example, denies what he calls the ‘Internal Aspects View’ in favor of the ‘Essentially Comparative View’ to reject the principle of transitivity of ‘better than.’ Still, there are various ways of coming to terms with the repugnant conclusion.⁴⁰ My point here is not that we should accept the repugnant conclusion. Regardless of who is right on this matter, rejecting the transitivity of ‘better than’ is clearly a radical step that challenges one of our most pervasive evaluative principles. My point is that counterexamples to formal intuitions yield paradoxes, whereas counterexamples to theoretical intuitions do not. Moreover, one plausible way to solve paradoxes generated by formal intuitions is to reject otherwise appealing concrete

⁴⁰ Huemer (2013, 332–5) claims that the ‘Money Pump Argument’ shows that intransitive betterness judgments give rise to paradoxical results. That is, in Huemer’s view, Temkin tries to solve a paradox by creating another paradox. Moreover, it may be true that the repugnant conclusion is *not* repugnant after all: why would it be repugnant to say that *Z* is better than *A*, considering that *Z* includes more of what is good? Since we compare the *total* quality of life, there may be nothing wrong with the claim that *Z* is better than *A*. There are also other ways to accept the repugnant conclusion. Tannsjö (2002, 347–9) denies that the repugnant conclusion is counterintuitive by appealing to the human capacity to adapt to new conditions. Arrhenius (2000, 251–9) accepts the repugnant conclusion by denying that there are any acceptable alternatives to the repugnant conclusion. For reasons why we should not try to avoid the repugnant conclusion, see also “What should we agree on about the repugnant conclusion?” *Utilitas* (2021, 1–5).

intuitions in favor of preserving formal ones. Formal intuitions are, therefore, helpful in assessing and – at least sometimes – revising our moral theories.⁴¹

The formal intuition (c) above seems to be a suitable candidate for a consequentialist formal intuition. Are there any deontological formal intuitions? One possible candidate is Kant's first formulation of the Categorical Imperative (also known as the universalizability principle): "*Act only in accordance with that maxim through which you can at the same time will that it become a universal law*" (GMS 4:421). In other words, your maxim (or the subjective principle that motivates you to perform a certain act) can qualify as a moral principle if and only if it does not lead to a contradiction when raised to the level of a universal law. For example, if everyone were to act on the subjective principle that "Lying is permissible if you can get away with it" with the same regularity of laws of nature, that would lead to a contradiction. This is because the assumption of truth-telling is a necessary condition for lying. Since nobody will assume that anyone is telling the truth, it is impossible for anyone to lie under that maxim. In such cases, we think that a particular consideration gives rise to a paradox about a formal intuition rather than giving up the intuition itself (the law of non-contradiction in Kant's case).⁴² Even if

⁴¹ Some examples of using formal intuitions in support of moral arguments are as follows: Huemer (2003, 147–71) employs formal intuitions to support an argument against welfare egalitarianism. The argument for the repugnant conclusion, as I mentioned above, is based mainly on formal intuitions, even though Parfit does not agree with the argument. Peter Unger (1996, 88–94) uses the formal intuition "If x and y are qualitatively identical in nonevaluative respects, then x and y are also morally indistinguishable" to bolster his argument for the claim that some of our intuitions about sacrificing people to produce greater benefit are unfounded.

⁴² I admit that declaring the first formulation of the Categorical Imperative to be a *deontological* formal intuition appears to be hasty. One might argue that the universalizability principle, on its face, is neutral between consequentialism and deontology. Surely, there is a 'universalizability' requirement built into the logic of moral judging, but that requirement does not seem to rule out consequentialism: the impossibility, logical or otherwise, of everyone's doing what I want to do is neither necessary nor sufficient to show that what I want to do is impermissible. So, the question is, how is it possible for a purely formal principle to produce substantive moral content? This, I

we accept that concrete and mid-level intuitions are unreliable due to the distortion created by nonmoral factors, we cannot reach the same conclusion for Kant's first formulation of the Categorical Imperative because it is a purely formal principle without any experiential content. Thus, it is not directly susceptible to evolutionary influences. Kant's universalizability principle itself may have nothing to do with our evolutionary history or emotions. It is devoid of any empirical content, yet it is the foundation of Kant's deontological moral theory. Emotions and evolutionary influences could come into play only when we try to derive concrete moral duties (e.g., "You ought not to commit suicide") from that formal principle.⁴³

It is possible that there exists both consequentialist and deontological formal intuitions that are safe from morally irrelevant factors and exceptions. Can we use these formal

think, is a legitimate concern; however, there are possible solutions. For instance, according to Oliver Sensen's (2014) interpretation of Kant's universalizability principle, the principle is recognizably moral because it represents the idea of *fairness*. That is, the universalizability principle forbids us from seeing ourselves as superior to others or making an *exception* for ourselves to the laws we regard as objectively necessary. This is Kant's *formal* principle. But we should resort to empirical sciences to discover these objective and necessary laws and to derive concrete duties from that formal principle. For example, we regard the law against killing objectively necessary because following it is necessary for us to be able to live at all. Or we regard the law of helping others as objectively necessary because it is a *necessary* means to an end *all* of us have, i.e., we will necessarily need help at some point in our lives as finite beings. It is to these laws that we should not make an exception. More on Sensen's interpretation in the third chapter.

⁴³ One could also say that Kant's universalizability principle does not rule out consequentialism because consequentialism could be a universal law in the sense that everyone could follow the rule "Always maximize utility." However, it is worth noting that what we are assessing is not *actions* themselves but *maxims*, namely the *reasons* behind actions. For example, take the consequentialist maxim "*Whenever X leads to a better consequence, do X.*" Substitute *X* with 'lying,' 'committing suicide,' 'committing murder,' 'not looking after one's children,' and so on, and suppose that these subjective principles become necessary and universal like laws of nature. If everyone were to act on these maxims with the same regularity of laws of nature, then either there would be *no* consequence to be improved (because everyone would die soon or older members of society would not be replaced), or it would ultimately become *impossible* for any of the mentioned actions to lead to any better consequence. Although I admit that more work has to be done to be able to claim that the universalizability principle and consequentialism are mutually exclusive, this is roughly how the universalizability principle could rule out consequentialism.

intuitions to make substantive normative judgments? This is the real difficulty that faces both consequentialists and deontologists when they try to prescind from the level of purely formal intuitions to particular cases. It seems that different moral theories can have different secure formal intuitions. However, once they try to fill the gap between formal intuitions and particular cases, their theories become vulnerable to exceptions and nonmoral factors. The problem that confronts all such theories is, thus, to find a way to translate formal intuitions into action-guiding general principles with content, without falling prey to exceptions and morally irrelevant biases. Although Greene successfully shows us the way in which *some* of our concrete and mid-level deontological intuitions *are* susceptible to distortions that arise from evolutionary forces in *some* moral dilemmas, his evidence does not seem to apply to *all* moral dilemmas, nor to (non-formal) theoretical and formal (theoretical) intuitions. Hence, he is not entitled to claim that consequentialism is normatively superior to deontology.

2.2.6 How to Acquire Theoretical Intuitions

I have argued that while our concrete and mid-level intuitions are susceptible to evolutionary influence, our theoretical intuitions (formal ones included) often are not directly susceptible to them. But are theoretical intuitions really “intuitions”? Or are they generalizations – convictions which we develop only *after* we *reflect on* particular cases? Since intuitions are “*immediate* intellectual grasps or appearances *prior to* reasoning,” it may seem to be wrong to classify them as intuitions. In this section, I describe how we could turn abstract moral theories into intuitions.

2.2.6.1 Two Systems

Actions arouse emotions. Due to our shared biological nature, we are inclined to react in particular ways to particular actions. These reactive attitudes are the products of our ‘*concrete intuitions*.’ For example, when we hear about an instance of incest, or someone beating his child, or someone cheating on their partner, our concrete intuitions make us feel that they are wrong, and we react accordingly. Then we use our capacity for abstraction (the capacity to see actions or objects as members of general categories) and find certain *kinds* of actions right or wrong (e.g., “Incest is wrong”). We call such reactions ‘*mid-level intuitions*.’ Mid-level intuitions are similar to what Haidt (2001, 828) calls “*post hoc* rationalizations:” They appear to originate from moral reasoning, but they are often merely expressions of our emotions (plus abstraction).⁴⁴ Then we systematically think about our concrete and mid-level intuitions and question their appropriateness, credibility, coherence with each other, and so on. We reflect on various circumstances we might find ourselves in and whether different conditions affect the rightness/wrongness of particular actions. We also try to see whether our concrete and mid-level intuitions check with our formal intuitions. After reflecting systematically on our concrete and mid-level intuitions, we reach generalizations or abstract moral theories. And when we can think of exceptions or counterexamples to our moral theories, we either reject them completely or revise them. This is roughly the way we exercise our capacity for autonomous moral reasoning. But it would be quite difficult – if not impossible – and

⁴⁴ Our capacity for abstraction differs from our capacity to make generalizations. While the former capacity does not involve any inference (e.g., mid-level intuitions), the latter capacity is exercised through drawing inferences (e.g., theoretical intuitions).

ineffective to stop what one is doing each time and repeat this process for each action. We certainly need context-sensitive intuitions to effectively guide our actions. Can our generalizations or moral theories convert to intuitions? And if so, how?

To explain how we turn moral theories into “immediate intellectual grasps or appearances,” we need to focus on Greene’s distinction between the cognitive and the emotional. Assuming a sharp distinction between the emotional and the (dispassionately) cognitive is too hasty, of course. A *purely* cognitive or emotional judgment could be an impossibility due to the extremely complex nature of the brain. What we call ‘the cognitive’ could involve emotions, or what we call ‘the emotional’ could involve cognitive elements.⁴⁵ However, it is plausible that we are *functionally* divided into two systems: emotional and cognitive systems. Each system is associated with a distinct region in the brain (Greene 2008, 40–1). We have an *unconscious body* that has been successfully adapted to its environment and thus knows what to do even when there is no input from our conscious mind. For example, we automatically and effortlessly judge foods containing sugar and fat tasty because in the environment in which our Pleistocene ancestors lived, high-calorie foods were rare, and they needed calories to survive. Evolved reactions of our unconscious body are fast, automatic, and effortless. On the other hand, we also have a *conscious “I”* that engages in conscious, slow, deliberate, and effortful reasoning that enables us to change our priorities and judgments. We know that in today’s fast-food culture, foods containing high amounts of sugar and fat may be dangerous because they could cause health problems such as obesity, diabetes, and heart disease. Thus, it comes as no surprise to us why many people today regard such foods as

⁴⁵ Cf. Greene 2008, Railton 2014.

unhealthy. More importantly, the input from our conscious “I” could even change our automatic reactions: we find many people today automatically and effortlessly react *in disgust* to some fatty and sugary food. My claim is that something similar may occur with our moral judgments. Over time, some of our moral judgments, which are generated by conscious reasoning, could transform into fast, automatic, and effortless reactions that allow us to grasp moral facts and distinctions immediately. These emotional reactions whet the eliminativist’s or the behaviorist’s appetite because our automatic reactions to moral cases suggests that these responses have been shaped exclusively by selective pressures. They think that they can safely claim that objective morality is an illusion by pointing out these automatic emotional reactions. However, they quickly disregard the possibility that those emotional reactions have a moral/rational content.

The distinction between the emotional and the cognitive has been popularized by Daniel Kahneman, who calls them “System 1” and “System 2,” respectively (Kahneman 2011, 20–4). *System 1* involves automatic, effortless, and unconscious reactions to particular cases. It is like an autopilot, a stranger within, that “operates automatically and quickly, with little or no effort and no sense of voluntary control” (Kahneman 2011, 20). *System 1* includes our innate, evolved, psychological dispositions that we share with other animals. It is our fast, effortless, and mostly unconscious mental system for jumping to conclusions to increase our chances of survival. Activities that are linked with *System 1* include “making a ‘disgust face’ when shown a horrible picture,” “understanding simple sentences,” “driving a car on an empty road,” and so on. By contrast, *System 2* “allocates attention to the effortful mental activities that demand it, including complex computations” (Kahneman 2011, 21). It is associated with deliberate, logical, and slow

reasoning that requires a lot of energy. *System 2* is an expert at solving problems, but it has limited capacity to exert effort. This is because we are not machines: it is difficult for us to maintain our focus for a long time and on different things at the same time. Activities that are linked with *System 2* include “searching memory to identify a surprising sound,” “parking in a narrow space,” “filling out a tax form,” “checking the validity of a complex logical argument,” and so on. Edward Slingerland (2014) uses the term “*hot cognition*” for *System 1* and “*cold cognition*” for *System 2*. He identifies hot cognition with “knowing how” and cold cognition with “knowing that” (Slingerland 2014, 29). He also claims that we need cold cognition, even though our conscious mind is costly and slow: (1) Hot cognition gets help from cold cognition when there are two or more competing desires or desires conflicting with our long-term goals, and (2) cold cognition helps us in managing our social lives by creating “complex modeling of other minds,” and “virtual representations of the internal thoughts,” which facilitate figuring out how to interact with others (Slingerland 2014, 59).

2.2.6.2 *Internalization of Reasons and Cultural Evolution*

Now I can explain how we turn our moral theories into intuitions. Our Pleistocene ancestors used to live in small groups of hunter-gatherers, and they interacted mainly with relatives or people well known to them. Under these conditions, they developed psychological adaptations to facilitate and maintain cooperation such as emotions (e.g., empathy, resentment, etc.), the ability to recognize and remember faces, detect cheaters, and so on. We share these hot cognition processes with some of nonhuman animals. The important question is, “how one particular primate, us, managed the abrupt transition

from our ancient hunter-gatherer lifestyle to the large-scale, urban way of life made possible by agriculture” (Slingerland 2014, 175). In other words, how it became possible for our *hominin* ancestors, equipped only with hot cognitions, to adapt to urban life? It is less likely that hot cognitions are responsible for this adaptation because the time scale is too short for the evolution of new and complex psychological dispositions. One idea is that the abrupt transition became possible because of the introduction of new, external social institutions, which are designed to keep hot cognitions in check. However, given that cold cognition has a limited capacity to exert effort, it doesn’t seem to be well-equipped for this task. Therefore, it is more likely that the transition to dense urban life “was managed not by consciously suppressing our tribal emotions but by using cold cognition to extend or redirect instincts through a process of *emotional education*” (ibid., 176; emphasis added). That is, instead of trying constantly to keep our pre-reflective intuitions in check, we create a set of shared values through the exercise of autonomous reasoning, and, more importantly, we *internalize* them: they become second nature to us.

The role of cultural evolution in achieving this is immense: cultural inculcation can train us to internalize the input from autonomous reasoning and react accordingly to particular cases so rapidly (even within a single generation) that biological evolution cannot be responsible for it.⁴⁶ The capacity for autonomous reasoning enables us to reflect on and evaluate our evolved psychological dispositions. We are also able to grasp objective facts about the world,⁴⁷ owing to this capacity. However, due to differing

⁴⁶ The fact that “[w]e have no evidence for biological change in brain size or structure since *Homo sapiens* appeared in the fossil record some fifty thousand years ago” suggests that “[a]ll that we have done since then [...] is the product of cultural evolution” (Gould 1996, 354).

⁴⁷ These facts include, but are not limited to, scientific facts, mathematical facts, and arguably moral facts and facts about beauty.

conditions we find ourselves in, judgments we reach vary hugely from one individual to another, or from one society to another. To make progress in any field requires eliminating false and distorted beliefs, being skeptical about implausible beliefs, and keeping true and plausible beliefs. Cultural evolution comes into play and enables us to disseminate ideas and judgments, eliminate false and distorted ones, and keep the good and promising ones. Culture can shape our beliefs and feelings over time by imposing *reasons* and thus facilitate *progress* in a given field. For example, no reasonable person today believes that the earth is flat or that the geocentric model of the universe is correct. Also, no reasonable person today believes that slavery or racial discrimination is good intrinsically. Cultural evolution seems to be responsible for the rapid transformation of ideas and the progress that transformation brings about.

Cultural evolution is “a distinctive human mode of *evolution* that has surpassed the biological mode because it is a more effective form of *adaptation*; it is faster than biological *evolution*, and it can be directed” (Ayala 2016, 252). We call the effect of culture on transforming our ideas and beliefs ‘cultural evolution’ due to the striking resemblance between evolution of ideas/beliefs and biological evolution. First, there is *heredity*: ideas/beliefs are transmitted from one individual to another or from one society to another. Second, there is *variation*: beliefs, ideas, and values differ among individuals and societies. For example, there are disagreements over scientific facts or moral issues such as abortion or euthanasia. And third, there is *differential reproduction*: some ideas/beliefs are transmitted more efficiently than others among individuals or groups.

Cultural evolution occurs much faster and more efficiently than biological evolution. There are two main reasons for that: (1) cultural heredity goes *horizontal* as well as

vertical, and (2) cultural mutations can be *directed* as opposed to the accidental nature of biological mutations. Biological heredity is Mendelian: it has only vertical transmission. This means that traits can only be transmitted from parents to children. Cultural heredity is, however, Lamarckian:⁴⁸ what we transmit to other members of society are not only the ideas/beliefs that we got from our parents but also ideas/beliefs that we have received from the whole human environment. We can transmit our ideas, beliefs, and values not only to our children but also to all humankind, irrespective of whether they are related to us. Thus, cultural evolution occurs much more efficiently than the biological one. Moreover, contrary to the accidental nature of biological mutations, cultural mutations can be directed through rational thinking. Biological mutations occur accidentally and only few of them are beneficial. On the other hand, we can consciously create cultural mutations according to our needs and reasons. We invent and discover ideas/things that we think are beneficial to us and we can disseminate them rapidly to all humankind in a generation. Thus, cultural evolution occurs much faster than biological evolution.

Just as some biological traits are favored by natural selection, so some ideas/beliefs are favored by *cultural selection*. Cultural ideas/beliefs compete for our attention and some of them are more successful in replicating themselves than others. What could be the reason for such a success? *Coercion* could be a possible reason. For example, after the Roman emperor Constantine the Great became a Christian in 312 CE, most of the

⁴⁸ In his 1809 book *Philosophie Zoologique*, Lamarck claims that we do not only transmit the genes we inherited from our parents when we have a recent mutation. We also transmit the attributes that we acquire as a result of that mutation. For example, according to Lamarckian heredity, when you go to the gym and develop big muscles, your children will be born with big muscles. Although this is not true for biological evolution, it is true for cultural evolution.

Western world suddenly became Christian.⁴⁹ *Psychological attractiveness* could be another reason. For a long time and in most cultures, people have believed that God exists. The reason for this success could be that it gives people security, love, and tranquility, i.e., the idea of God is psychologically appealing. However, some of the cultural ideas/beliefs' survival and reproductive success is independent of biological fitness. We often opt for certain beliefs, not because they are psychologically attractive but because they *track the truth*. For example, the reason why no reasonable person believes that the geocentric model of the universe is correct is *not* because the belief in question is psychologically unappealing but because it does not track the truth. The capacity for autonomous reasoning creates a gap between biological and cultural evolution by enabling us to create cultural mutations that track the truth, regardless of their biological advantages.⁵⁰

Cultural evolution enables us to develop ideas and find solutions to long-term problems much faster and more efficiently than biological evolution, because cultural heredity is *horizontal*. However, our ability to use cultural information is restricted because cold cognition is slow, effortful, costly, and limited. The solution is *domestication* or *internalization*: just as we can deliberately change the behavior of

⁴⁹ Cf. Jones 1948, 79–83.

⁵⁰ Many behaviors have been moralized and amoralized in short periods of time. For instance, smoking has been moralized recently, while homosexuality, divorce, marijuana use, and atheism have been amoralized (cf. Pinker 2002, 275). If we have the capacity to reason autonomously, then the way we assess our evolved dispositions could be similar to the way we assess scientific beliefs. That is, it is possible that we have been systematically trying to see whether our pre-reflective moral intuitions correspond to objective moral facts. And it is possible that ideas that do not track the moral truth are selected against during cultural selection. If this is in fact what is happening, then we can regard some of the moralization/amoralization cases as instances of moral progress.

animals and plants in line with our needs, so “the conscious mind can acquire new, desirable goals and then download them onto the unconscious self, where they can then be turned into habits and implemented without the need for constant monitoring” (Slingerland 2014, 65). For example, when beginner drivers learn how to drive, prefrontal regions in their brains become much more active because they constantly try to maintain their focus to understand the traffic rules, how to turn the wheel, how to press on the pedals, and so on. However, after certain amount of practice, the nervous driver starts talking and making jokes while driving because their brain activity decreases. As they *internalize* driving skills, their conscious mind or cold cognition becomes less active. Likewise, beginner chess players have to think a lot about how pieces move, how to devise strategies to win, and so on. But over time, thanks to the human capacity to recognize patterns, they develop intuitions that allow them to make quick and efficient decisions without having to think or calculate for too long (especially in bullet chess games, where each player is given only a minute in total, players mostly play by their ‘chess intuitions’ instead of active thinking). In both examples, cold cognitions are transformed into hot cognitions over time.

In similar fashion, it could be that we turn our abstract moral theories to theoretical intuitions over time. Although we arrive at moral theories through conscious deliberation, we may later turn them into hot cognitions. Our theoretical intuitions may help us grasp moral facts and distinctions and react to moral cases *immediately*, without thinking much about them, like a chess grandmaster playing a bullet game. Admittedly, chess intuitions might go wrong because they are not perfectly reliable in grasping chess facts. Chess grandmasters usually analyze games they played to improve their skills and to learn from

their mistakes. And once a better move or strategy is discovered, the chess grandmaster revises or gives up their chess intuitions. Similarly, our theoretical intuitions are *not perfectly reliable* in grasping moral facts and distinctions. Therefore, when a counterexample is raised, we question the credibility of our theoretical intuitions; we either revise them, or we give them up.

2.2.6.3 Meme Theory

Before I move on to the question of how theoretical moral intuitions affect our reactions to particular moral cases, I would like to address an objection to the proposed account, namely Dawkins's meme theory.⁵¹ Dawkins's meme theory attempts to account for cultural evolution by associating it with evolved human psychology. He thereby attempts to close the explanatory gap between genotype and phenotype. The basic idea is that genes are not the only replicators in nature. There are units in nature that share the same defining features with genes.⁵² Dawkins calls these units 'memes.' Examples of memes are religious or scientific ideas, melodies, artworks, catch-phrases, clothing trends, and so on. Memes replicate themselves by imitation, they constitute culture, and they establish their own course of evolution that is independent of genetic programming. What is more, since cultural evolution occurs much faster than the biological one, it is at the forefront of human evolution. Nevertheless, memes are still affected by genes in the sense that the survival of a meme depends on the level of its *psychological appeal* and genes make

⁵¹ Dawkins 2006, 189–201.

⁵² The defining features of replicators are longevity, fecundity, and copying-fidelity (Dawkins 2006, 15–8, 194).

some memes more appealing: “Psychological appeal means appeal to brains, and brains are shaped by natural selection of genes in gene-pools. They want to find some way in which having a brain like that improves gene survival” (Dawkins 2006, 193). Dawkins’s background hypothesis is that human psychological and motivational constitution has been shaped by natural selection. We are inclined to imitate/replicate cultural ideas that have a stronger psychological appeal. And the reason we find some ideas more attractive is that they carry with it a higher survival value. The idea of God is a prominent example. It follows that cultural ideas are subject to genetic influence: genes render some ideas more attractive due to their higher survival value.

One question one could raise is, can natural selection explain each of the billions of distinct thoughts or ideas, given that human genotypes are more or less similar? Why are we so different in what we think when our basic psychological and motivational constitution appears to be uniform from a biological point of view? It is likely that something other than human genotype must explain the enormous heterogeneity of human thought, which suggests that our adoption of some memes/thoughts/ideas may be independent of their survival value. For instance, why would we find the widespread meme “God does not exist” psychologically appealing? Is it conducive to survival? Or think about contraception: Why would too many people support such an idea if psychological appeal were the only criteria for an idea’s survival and reproductive success?

Furthermore, Dawkins seems to disregard the role of reasons in accepting a thought or an idea. The success of many memes depends not on their psychological attractiveness but on whether they track the truth. An obvious example is scientific ideas. People

generally adopt scientific ideas not because they are psychologically attractive or because they increase the chances of surviving but because they think they have *good reasons* to believe that the ideas they are reflecting on *are true*. The crucial question here is, do we have good reasons to think that ideas about morality are different from scientific ideas in this respect? And if yes, to what extent they are different from one another? The debunker is not justified in her assumption that there is a pervasive evolutionary influence on the content of our moral beliefs unless she gives satisfactory answers to these questions and espouse the Dawkinsian picture, and unless she plausibly rejects the role of cultural evolution and autonomous reasoning, processes that may occur independently of biological evolution in shaping the content of our moral beliefs.

Recall that the burden of proof must be on the debunker (2.1.2). That is, the debunker needs to provide a good reason that we are probably mistaken in our evaluative/moral beliefs. To do that, she must provide evidence of error supported by reliable scientific data. This means that the debunker needs a strong empirical premise to present a strong argument. *Proving* that there is a pervasive evolutionary influence on the content of our beliefs, without a doubt, would make a strong empirical premise. However, merely *asserting* a pervasive influence would amount to an outright rejection of the influence of autonomous reasoning and cultural evolution, which appears to be an illegitimate move due to considerations raised above.

2.2.6.4 *Our Reactions to Moral Cases*

If the proposed account is true, how does it explain our reactions to moral cases? Once we turn our moral theories into hot cognitions, they start informing our concrete

intuitions and affect our reactions to particular cases. Greene may be correct in asserting that the nonmoral factor of personal force distorts most people's deontological judgments when they react differently to trolley and footbridge dilemmas. However, this may be because the effect of personal force becomes more distorting in unlikely, unrelatable, and unrepresentative situations like trolley-type cases. In a relatable case, such as the Bus dilemma, the distorting power of personal force becomes less active, and thus many people can make a moral distinction between 'harming an innocent person' and 'directing a public threat to a lesser harm' more easily. Furthermore, the Bus dilemma makes people revise/reject the deontological theoretical intuition that "It is wrong to harm an innocent person, no matter what the consequences are" and adopt instead the consequentialist theoretical intuition that "It is permissible to harm an innocent person, if your action saves more innocent people." This consequentialist theoretical intuition informs people's intuitions about the Bus case (concrete intuitions) and enables them to grasp the moral fact that "It is (morally) permissible to push the corpulent gentleman in the Bus case."

Similarly, in the Sadistic Case, people revise/reject the consequentialist theoretical intuition that "An action is good, if it increases the overall happiness" and adopt instead the deontological theoretical intuition that "It is wrong to harm an innocent person for pleasure, even if it increases the overall happiness." This deontological theoretical intuition informs people's intuitions about the Sadistic Case (concrete intuitions) and enables them to grasp the moral fact that "It is wrong to torture the homeless person for pleasure in the Sadistic Case." Note that the reason we prefer a deontological intuition to a consequentialist one in the Sadistic Case is not that the effect of personal force distorts our judgments. As the Bus dilemma revealed, when a moral dilemma is set in a likely and

reliable context with more representative experience, our judgments are less likely to be influenced by nonmoral/evolutionary factors. Our negative reaction to the Sadistic Case is, thus, more likely to be the result of the cooperation between autonomous moral reasoning and cultural evolution.

2.2.7 Conclusion

I have argued that (1) the effect of personal force on our moral judgments fails to provide an adequate explanation of the different reactions to some morally indistinguishable dilemmas, that (2) moral dilemmas that are set in unreliable and unrepresentative contexts may prevent people from grasping objective moral facts, that (3) the theory I support offers a better explanation of our consequentialist and deontological responses to particular cases than Greene's theory since both consequentialist *and* deontological theoretical intuitions seem to be immune to nonmoral biases, and that (3) Greene is not justified in his claim that deontology is normatively inferior to consequentialism.

I have also argued that the best way to respond to the epistemic challenge posed by evolutionary debunking arguments is to focus on their empirical premise and show that it is *not* as strong as the debunker thinks it is. EDAs are essentially *inductive* arguments with a *probabilistic* conclusion, and the more set of beliefs an EDA calls into question the harder it becomes to provide a complete evolutionary origins story. Asserting a widespread evolutionary influence on our evaluative/moral beliefs amounts to espousing a particular metaethical interpretation of the empirical data provided by evolutionary biology: it echoes the eliminativist or behavioristic conception of morality that rejects objective morality. This means that the empirical premise commits the debunker to the

assumption that evaluative/moral realism is false. This is a philosophically illegitimate move considering that the skeptical conclusion of an EDA maintains that evaluative/moral realism is an implausible metaethical position.

To reveal the weakness of the empirical premise, I have focused on the effects of autonomous (gene-independent) moral reasoning and cultural evolution on the content of our evaluative/moral beliefs. I have argued that we have theoretical and formal moral intuitions, which are more likely to be immune to distortions stemming from our evolutionary past than our concrete and mid-level moral intuitions. Owing to the processes of autonomous moral reasoning and cultural evolution, we may be able to turn our moral theories and reasons into theoretical intuitions, which may inform our intuitions about or reactions to particular moral cases or dilemmas by making it possible for us to grasp moral facts and to make important moral distinctions. Thus, EDAs are not strong enough to undermine moral realism or non-naturalist yet stance-dependent accounts such as Kantian constitutivism.

CHAPTER 3: KANTIAN CONSTITUTIVISM

In the first chapter, I set my aim, which is to find a plausible alternative to moral realism. My claim is that Parfitian moral ontology supported with a Kantian constitutivist origins story could be that alternative. I also claim that this position could be seen as a *stance-dependent* view, so it is not realism, but it is a form of constructivism. But I have not explained why this view could be seen as a neglected and plausible alternative to realism. In this chapter, I explain what constructivism is, what kinds of constructivism there are, and why I think only the Kantian constructivism, in the form of constitutivism, seems to capture objectivity and categorical normativity of morality, and thereby poses an alternative to moral realism. I also discuss the advantages and disadvantages of such a view and see if it is really a plausible alternative; that is, if it has less problems than realism.

Let's start with Shafer-Landau's (2003, 15–8) metaphysical classification. According to him, there are three theories of the metaphysical source of morality: (1) *nihilism*: there is no morality; there is no stance-independent moral goodness that makes our moral judgments true; (2) *moral constructivism*: morality is a human construct; that is, the truth-makers of moral judgments are our contingent choices or agreements; (3) *moral realism*: morality is independent of any standpoint we may take; that is, moral statements are made true independently of desires, preferences, conventions, agreements, and so on. He then subdivides constructivism into three distinct views: (2.1) subjectivism: morality is

constructed by an individual; (2.2) relativism: morality is constructed by two or more people; (2.3) ideal standpoint theories: morality is constructed by decisions made under *ideal* conditions of choice such as Rawls's veil of ignorance. This classification is not exhaustive because it rules out Kantian constitutivism, according to which morality is constitutive of practical reason. In the following, I explain why Kantian constitutivism – especially in the form of transcendental constitutivism – is a neglected alternative to moral realism.

In the first chapter, I argue that moral realism has big metaphysical and epistemological problems. Realism takes moral phenomenology seriously and captures the common-sense conception of morality as being objective and categorically normative. This is the advantage of realism. But this advantage comes at the cost of postulating further non-natural entities, waiting 'out there' to be discovered by us (or being tied to human rationality as an absolute inner value), and we have seen how problematic such postulations could be. Postulating such entities leaves us with a possibly unsolvable mystery: it is quite difficult to make sense of how realism can plausibly account for supervenience and of how moral knowledge is possible under this view.

What is more, we should not take for granted the realist explanation of normativity. It may be true that moral normativity can only be captured by a non-naturalist account. But realism takes normativity to be generated out of the *third-person* perspective: there are impersonal objective facts 'out there' that are independent of any human standpoint (or these facts are created by an absolute inner value that is tied to human rationality). Claiming that morality follows from a third-person standpoint raises an important question: what obligates one to follow these stance-independent impersonal facts?

Perceptual characterization of moral knowledge and the third-person perspective not only create big metaphysical and epistemological problems, but they also impose an explanatory burden on realism: it is not easy for realism to explain why morality is *authoritative*, i.e., why is it binding on us. Let's grant that someone has intuited the goodness of a certain action. Why perform that action when one is sure that it involves an unwelcome cost?¹

I also claim in the first chapter that evolutionary explanations of our moral intuitions threaten to undermine our common-sense conception of morality as objective and categorically normative. If our sense of moral ought is just the product of evolutionary processes, i.e., if the sole function of our moral sense is to promote and maintain cooperation, then it seems, at least on the face of it, that we have strong inductive reason to believe that the common-sense conception is merely an illusion, i.e., that there is no objective morality. This is the nihilism option adopted by Mackie and Joyce.² I assert in 1.4.3 that if the only way to capture the indispensable features of morality were to ascribe irreducible moral properties to actions, persons, or character traits, then perhaps it would not be irrational to reject the existence of morality altogether. But this would still leave us with a highly unappealing and counterintuitive conclusion because, if true, nihilism entails that all our moral judgments are false. That is, if morality does not exist, then statements such as "Torturing infants for fun is wrong" or "Genocide is wrong" cannot be true, which sounds highly implausible. My aim in this chapter is to show that positing non-natural moral entities is not the only way to capture the essential features of morality.

¹ Korsgaard 2003, 110–2.

² Of course, there are other ways to reach the nihilistic conclusion such as the argument from queerness and the argument from relativity.

Moral nihilism or error theory is unattractive because it is highly counterintuitive to claim that we are all wrong about our moral claims. But Mackie's argument from queerness gives us reason to be parsimonious about our ontology. It is true that the mere claim that objective values would be different from anything else in the universe does not give us sufficient reason to deny objective morality. On the other hand, as the discussions about supervenience and intuitionism show, positing non-natural moral entities seems to inflate our ontology for no clear reason at all. Moral realism could be right in its claim that morality is objective, and nihilism could be right in its claim that the realist metaphysics is problematic. So, it seems that a plausible approach would be to be as parsimonious as possible in our moral ontology, while still being able to give a satisfactory account of objectivity and categorical normativity. This is the reason I explore constructivism and see whether it can give us what we need.

In 3.1, I talk about the general constructivist project and different forms of constructivism. I then show why idealized stance constructivism and Humean constitutivism cannot give us what we need, in 3.2 and 3.3, respectively. In 3.4, I explore two forms of Kantian constitutivism: Korsgaard's Kantian constitutivism (3.4.2) and Sensen's transcendental constitutivism (3.4.4). I claim that transcendental constitutivism has a parsimonious moral ontology and can potentially account for moral objectivity and categorical normativity, thus being a neglected and a plausible alternative to moral realism.

I then move on to discuss possible objections to transcendental constitutivism in 3.5. First, I discuss the objection that transcendental constitutivism is as mysterious as moral realism because it regards *pure* reason as the source of morality and thus it appeals to a

noumenal realm (3.5.1). I claim that transcendental constitutivism does not appeal to a noumenal realm due to the distinction between independence from *causation* in nature and independence from *existence* in nature. That is, the fact that pure reason generates the moral law spontaneously does not entail that pure reason is ontologically independent of the natural world. Rather, we can conceive of pure reason as residing in nature as an emergent and unalterable structure of thinking. I also mention possible allies of this view such as Chomskyan linguistics, Mikhail's universal moral grammar, functionalism in the philosophy of mind, and Fodor's modularity of mind. I consider them as allies because they all place function of reason or mind at the center.

In 3.5.2, I discuss the bootstrapping (or emptiness) objection, according to which it is implausible to bootstrap substantive reasons into existence from a merely formal moral law. In other words, the objection claims that we cannot derive substantive morality out of thin air. I concede that a purely formal law, considered in and of itself, is empty. We really cannot derive determinate moral content from a purely formal law, if it is the only source we have. However, I argue that, if our reason draws inferences from empirically identifiable universal human ends to determine substantive moral content and makes the necessary means to these universal ends binding through the formal moral law, namely the Categorical Imperative, then the objection may fail.

Finally, in 3.5.3, I revisit the evolutionary challenge to non-naturalist accounts of moral objectivity. Evolutionary considerations, *prima facie*, pose a serious threat to transcendental constitutivism. Nevertheless, transcendental constitutivism is compatible with the idea that our sense of moral objectivity and categorical normativity have promoted survival and reproductive success. According to transcendental constitutivism,

if we could create a free being, i.e., a being with a mind that has a certain level of complexity, out of nothing, that being would be under the moral law and would have a sense of moral ought. This is against evolutionary explanations such as Joyce's, according to which internalization of a moral sense requires evolutionary history. However, I claim that transcendental constitutivism does not entail that individuals can create morality (as we know it) or moral behavior in isolation from each other. On the contrary, morality is realizable for creatures like us through the development of a social community. That is, if a being possesses a reason that is free, she is immediately under the moral law (or the Categorical Imperative). The content of the moral law "does not alter;" it is independent of our evolutionary history. However, to be able to derive specific moral rules and to reach moral judgments, we need evolutionary history. Transcendental constitutivism does not dismiss the importance of our evolutionary history in explaining the emergence of human morality as we know it and some of the content of our moral judgments. Thus, evolutionary considerations do not pose a serious threat to the main claims of transcendental constitutivism.

3.1 The Constructivist Project

There are two thoughts that motivate the constructivist project. First, there are genuine normative or moral truths. And second, these truths are not part of the fabric of a stance-independent reality. That is, moral properties do not exist 'out there,' nor they are tied to our rationality as an absolute inner value, as different versions of realism have it. As we have seen, if the source of moral truths is stance-independent moral properties, we would be isolated from those truths in metaphysically and epistemologically problematic ways.

To solve these problems, constructivism asserts that there is an intimate connection between moral truth and human standpoint.

Constructivism is a middle ground between realism and nihilism. Instead of arguing for stance-independent irreducible moral facts, constructivism asserts that moral facts are actual or ideal *human constructs*: “[t]he central idea behind constructivism is that moral values and moral norms are not discovered, or revealed to us as if by the gods, but rather *constructed* by human agents for specific purposes” (Bagnoli 2013, 1). Constructivism, like realism, is an umbrella term that encompasses a range of views. Just as the nature of stance-independent moral properties and facts have been understood in different ways by different moral realists, moral constructivists have various options available to them. The constructed moral reality could be the result of (a) *actual* preferences, choices, or agreements (subjectivism and relativism), (b) decisions made under *ideal* conditions of choice (idealized stance constructivism), or (c) what is *constitutive* of practical reason, agency, or the attitude of valuing (constitutivism).

There are certain advantages of the constructivist project in general. First, if morality is essentially a human construct, then we can free our explanation of objective moral truths from redundant metaphysical baggage. That is, we do not need to explain supervenience because if morality is a human construct, it can be given a naturalistic explanation.

Second, if morality is constructed by our cognitive activities, it may be relatively easy for us to have moral knowledge. It is easier to make sense of the way we know moral truths because it is *our* cognitive activities that create the moral standards. Just as we are

familiar with the rules we create in other areas, such as games and etiquette, we are familiar with the moral rules we create.

Third, certain forms of constructivism (especially the ones that adopt the first-person perspective, such as subjectivism and constitutivism) may explain authority or bindingness of morality more easily than other accounts (other forms of constructivism and realism). If morality is the product of *my own* reason or desires, then it may be easier to make sense of why morality is binding on me. On second-person (relativism, idealized stance constructivism) and third-person (realism) theories it is harder to explain why one should be moral. Why should I be motivated by the benefit that will result from complying with group preferences or agreements rather than by the benefit I will get from doing what I want and from dismissing society's rules? If the only reason to be moral is to advance self-interest, then I seem to have the reason to defect out of self-interest when the conditions are appropriate.

Some versions of constructivism are more ambitious than others. Constructivist theories that ground the standard of correctness for our moral judgments in our *actual* attitudes are the least ambitious ones in terms of capturing common-sense morality. These theories reject moral objectivity outright by taking moral truth to be dependent on actual desires or preferences of individuals (subjectivism) or of groups (relativism). Due to the aim of the dissertation, I will confine myself to brief remarks. Although subjectivism and relativism make no assumptions other than identifying morality with individual or group preferences, they redefine morality. In fact, their definition of morality is more radical than Railton's "reforming definition" of morality because they reduce morality to *actual* attitudes. According to subjectivism, the content of morality

changes whenever one has a different desire, and according to relativism, the content of morality changes whenever two or more people agree on something. Making morality dependent on actual attitudes amounts to ignoring our first-person experience of morality altogether. Think about bad conscience. Sometimes you think you did what you wanted to do but at the same time you think your action was morally wrong. When we talk about morality, we do not just talk about our desires; rather, we talk about something that *restricts* our desires. So, subjectivism and relativism simply change the subject. Worse yet, if morality is a function of actual preferences, then there are countless many moral (!) systems some of which allow murder, torture, rape, genocide, slavery, and so on. For instance, if subjectivism is true, it seems that a serial killer has a right to abduct and kill people provided that *he* believes that he has a right to abduct and kill people. Thus, under such a view, it seems that it is possible to justify *any* voluntary action regardless of its content.

The other two types of constructivism, namely idealized stance constructivism and constitutivism, are more ambitious in terms of capturing common sense morality because they attempt to describe a moral reality which is constructed yet objective. That is, the aim of such theories is to capture the common sense understanding of morality as objective and binding, while avoiding the problems associated with realism by offering a *stance-dependent* account.

According to idealized stance constructivism, if moral truth consists in principles that “no one could reasonably reject as a basis for informed, unforced general agreement” (Scanlon 1998, 153) or principles that can be acted on by *all* agents with “minimal rationality and indeterminate mutual independence,” (O’Neill 1989, 215) then morality is

not constructed by *actual* preferences, choices, or agreements. Rather, moral standards and facts arise from decisions made under *ideal* conditions of choice, namely the conditions of full rationality. The idea (and hope) is that under these ideal circumstances, our responses and desires would converge. This is their sense of objectivity.

According to constitutivism, moral truth consists in principles that are *constitutive* of practical reason, agency, or the attitude of valuing. This means morality is even independent of decisions made under ideal conditions of choice. Moral standards are laid out preconsciously and nonvoluntarily, i.e., they are independent of actual and ideal attitudes, but they are dependent on or constitutive of human reason (Sensen 2013), human action (Korsgaard 2009), or the attitude of valuing (Street 2010), i.e., they are *stance-dependent*. It may be, *prima facie*, easier on this first-person perspective view to explain the authority of morality: (a) morality is independent of actual or ideal attitudes; (b) morality follows from the structure of *our* reason, *our* action, or *our* attitude of valuing. We should be motivated to comply with moral rules because they spring from *our* nature. Although a fully convincing answer to ‘Why be moral?’ might be impossible,³ constitutivism may be better suited to account for the authority or bindingness of morality than some other forms of constructivism and realism.

3.2 Idealized Stance Constructivism

Idealized stance constructivism, understood broadly, claims that normative or moral truth is a complex function of a human stance via some form of idealization. The idea is that

³ This is because we cannot convince each skeptic. There will always be someone who will say, “So what?”

normative or moral facts are constructed by our decisions which are made under suitably constrained conditions of choice. This view is different from moral realism because, unlike realism, it takes normative or moral principles to be “constructed” rather than “discovered.” That is, such principles are not established through stance-independent moral facts that are “distinct from how we conceive of ourselves.”⁴

We can talk about two types of idealized stance constructivism: (1) *procedural constructivism*, and (2) *ideal observer theories*. According to procedural constructivism, “moral objectivity is to be understood in terms of a suitably constructed social point of view that all can accept.”⁵ That is, procedural constructivism takes moral truths to be constructed from an idealized decision procedure. To name a few examples, Rawls’ (1971) political constructivism, O’Neill’s (1989) Kantian constructivism, Scanlon’s (1998) contractualism, and Copp’s (2007) society-based constructivism associate moral truth with some kind of constructive procedure. Ideal observer theorists, such as Firth (1952), Railton (1986), and Smith (1994), also take morality to be constructed from an idealized human stance. The difference is that ideal observer theories define moral truth as a function of *desires* or *responses* of a fully informed and coherent agent (cf. 1.4.2). Central to both types of idealized stance constructivism, as the name suggests, is the idea that certain states of affairs constitute the ideal conditions for determining normative or moral facts and that no normative or moral fact exists independently of the decisions or responses made under such ideal conditions.

⁴ Rawls 1980, 519.

⁵ Ibid.

Idealized stance constructivism is ambitious in that it aims at a moral reality which is constructed yet objective. As stated above, the crucial claim of idealized stance constructivism is that no normative or moral principle is independent of how we conceive of ourselves. According to Rawls, we conceive of ourselves as “free and equal, [...] capable of acting both reasonably and rationally.”⁶ Our self-conception as free, equal, and rational beings paves the way for the possibility of arriving at *objective* moral principles. If we can find a reasonable ground for agreement, i.e., if we can come up with principles that all reasonable agents can accept, then the normative or moral facts that we build upon those principles, and the principles themselves, will be objective. The hope is that we, qua rational beings, will all agree on certain principles if the ideal conditions of choice are met. Likewise, ideal observer theorists have the hope that we will *desire* the same things or *respond* similarly to the various situations we might find ourselves in if the conditions of rationality are met: “the existence of reasons presupposes that under conditions of full rationality we would all have the same desires about what we are to do in the various circumstances we might face.”⁷ Idealized stance constructivists thus are after normative or moral facts that can be endorsed by *all* rational beings rather than by a certain group of people.

3.2.1 Idealized Stance Constructivism and Objectivity

If the hopes of idealized stance constructivism are realistic and reasonable, then this is good news for the supporters of moral objectivity who are dissatisfied with moral realism

⁶ Ibid., 518.

⁷ Smith 1994, 198. Cf. 1.4.2.

due to the epistemic and ontological costs associated with it. According to idealized stance constructivism, morality is objective in the sense that we will *all* agree on certain moral principles under ideal conditions of choice. This sense of objectivity rules out contingent attitudes because it rejects the idea that morality is constructed by *actual* preferences, choices, or agreements.

The question is, is this sort of objectivity is satisfactory? A satisfactory account of objectivity, at least for the purposes of this dissertation, is the one that gives us grounds for thinking that some of our ordinary moral statements can be *properly justified*. More specifically, if we are to properly justify our moral beliefs, we must show that the truth-maker of our beliefs is not individual or societal preferences. Otherwise, we would merely be observing contingent desires of individuals or societies and wrongly calling our observation ‘morality.’ If racism is wrong because some group of people decided so, then racism will be right if they change their opinion. Alternatively, racism could be wrong for some individuals and acceptable for others, depending on what they think or feel. Surely, we don’t want to say that, considering our commitment to morality’s objectivity (cf. 1.1).

Furthermore, if we are to properly justify our moral beliefs, we must also show that the truth-maker of our beliefs is *not* our biological nature. We all want to survive and reproduce as a species. If we cannot survive and reproduce, our species will cease to exist. There could be certain ways to ensure survival and reproductive success. Rules against murder or stealing, looking after our children, and truth-telling could all be necessary for *any* society to exist at all (cf. 1.4.4). Interestingly, many of our moral judgments are in fact about the wrongness of unjustified killing, special obligations to our

children or family, and the importance of honesty. Should we now say that moral facts just are the results of instrumental reasoning about what makes us good social cooperators?

The answer, I think, is no. First, desires are not good grounds for moral objectivity because they are subject to change. If we all had an intrinsic desire to harm as many people as possible instead of an instinct to survive and reproduce, then assault, lying, and manipulation would be morally right because they are efficient ways to harm someone. If we ground morality on something contingent such as desires, then morality becomes subject to change. If we had a different nature, then racism or torture, along with many other actions that we deem morally wrong, could have been morally right. Again, we don't want to say that.

Second, it makes sense to claim that evolutionary processes and our ultimate desire to survive and reproduce not only push some of our beliefs toward moral truth, but they also *distort* some of our moral judgments. It is true that rules against murder, looking after our children, and truth-telling make us good social cooperators, and we also regard many actions that conform to such values as morally right. Nevertheless, it seems that many of our evolved psychological dispositions have pushed our beliefs away from moral truth. Rigid gender roles, racial prejudices, disgust towards homosexuality, sexual taboos, notions of impurity, and hierarchical authority relations could be the byproducts of our ultimate desire to survive and reproduce along with human culture. Should we say that such practices, desires, or inclinations generate moral truth? Once again, we don't want to say that. A satisfactory account of moral objectivity, then, does not base morality on *any* desire, be it universal or not.

Can idealized stance constructivism provide us with a satisfactory account of objectivity while avoiding the epistemic and metaphysical problems associated with moral realism? It is in principle possible that the ideal conditions of choice are desire-independent. So, idealized stance constructivism could potentially give us the desired account of moral objectivity. But caution is needed here. For idealized stance constructivism to be a metaethical alternative to moral realism it must meet *two* conditions. First, we must be able to say that it is a *metaethical* view rather than a normative one. That is, it must make claims about the status or nature of morality without telling us, explicitly or implicitly, what *is* morally right or wrong. Second, it must give us a satisfactory account of moral objectivity.

Idealized stance constructivism fails to meet both conditions. First, many examples of idealized stance constructivism in the literature attach substantive (unconstructed) normative judgments or moral values to the idealized procedure. Second, idealized stance constructivism is question-begging as a metaethical option because it is subject to a Euthyphro-style dilemma: if morality is constructed under appropriately specified conditions of choice, then either these idealized conditions have a moral quality or not. If they do not have a moral quality, then we should not expect moral conclusions to follow (Shafer-Landau 2003). Actions that are usually considered as immoral could easily follow from such a procedure. If they *do* have a moral quality, however, then explaining these ideal conditions by appealing to stance-dependent moral facts is question-begging. So, idealized stance constructivism must either appeal to unconstructed, stance-independent moral norms and become a realist view, or the view becomes question-begging. And third, since idealized stance constructivism derives moral conclusions from

arbitrary (not properly justified) normative standpoints, the output of the procedure can at most be intersubjectively, as opposed to objectively, valid. This view can have its own merits, but this is not the kind of objectivity that this dissertation is after.

3.2.2 Normative Restrictedness

Idealized stance constructivism is not an alternative to any metaethical view because it attaches *substantive normative facts* to the procedure of construction. For example, according to Scanlon's contractualism, the procedure that creates normative truths is shaped by independent facts about good reasons to reject principles. The facts about what principles a group of rational beings would find non-rejectable are not themselves constructed by the procedure. Rather, they exist prior to the procedure and guide agents' deliberation in determining the relevant normative truths.⁸ Such a procedure can yield conclusions only about *some part* of normative domain; thus, such views have been labeled 'local' or 'restricted' constructivism.⁹ Similarly, Rawls' procedure is characterized by independent facts about the fairness of a hypothetical choice situation and facts about our self-conception as free, rational, and equal beings.¹⁰

Idealized stance constructivism does not qualify as a distinct metaethical position because procedural constructivists like Scanlon and Rawls do not specify the truth-maker of the normative or evaluative starting points that ground the constructive process. Is the

⁸ Scanlon 2014, 96–104.

⁹ Enoch (2009) and Bratman (2012) use the term 'local,' while Street (2008, 2010) uses the term 'restricted.'

¹⁰ Rawls 1980, 516–8.

value of the non-rejectability of principles determined by something factual or formal? Is the value of fairness constructed by contingent agreements, or does it exist independently of any human standpoint? Procedural constructivists choose to remain silent about such metaethical questions. Similarly, ideal observer theorists do not specify what makes full information and coherence valuable or morally significant. Accordingly, one can in principle be a realist, constructivist, expressivist, or error theorist about the nature of the normative facts idealized stance constructivists resort to in characterizing the procedure of construction or the ideal conditions of choice.¹¹ Since metaethical questions are set aside, adopting different conceptions of the constructive procedure simply means having a substantive *normative* disagreement.

Idealized stance constructivism cannot provide us with a satisfactory sense of objectivity because it derives moral conclusions from arbitrarily assumed starting points. For instance, Rawls' procedure involves the values of Western pluralistic democracy, and it excludes the values of non-democrats and marginalized people (O'Neill 2003, 359). Moreover, even if all people in a society agreed on the truth of such normative starting points, their moral judgments could at most be *intersubjectively valid* because their reasoning would only be instrumental. This is because if idealized stance constructivists choose not to specify the truth-maker of the normative input of the procedure of construction or ideal circumstances, then, for them, the most important function of practical reasoning must be that it provides us with the most efficient and sensible ways to satisfy *desires* every finite, mutually interdependent, rational, and biological being is supposed to have. Morality, then, must be reduced to the proper satisfaction of desires

¹¹ Cf. Street 2010, 368.

such as the desire to have a fair system of cooperation (Rawls 1980, 518), the desire to be able to justify ourselves to each other (Scanlon 1998, 202), or the desire to live in a peaceful, secure, and harmonious society (O'Neill 1989, 3–50).

3.2.3 Euthyphro Dilemma

Idealized stance constructivism is committed to the basic constructivist idea that no normative or moral truth exists independently of the practical standpoint. However, it does not specify the truth-maker of the conditions optimal for practical reasoning. For example, can a proceduralist constructivist appeal to the constructive procedure in accounting for the normative input of the procedure? Such a move subjects the view to a Euthyphro-style dilemma regarding the order of determination.¹² If normative truth is constructed under conditions optimal for making normative choices, then either these conditions involve normative constraints, or they do not. If the ideal conditions of choice do not have a normative (or moral) quality, then the standards or conditions constituting the practical standpoint must be arbitrary and no normative conclusion should be expected. If the ideal conditions do involve normative constraints, then it becomes impossible to justify them in a non-question-begging way, i.e., by appealing to the procedure of construction itself.

Idealized stance constructivists must explain how the normative constraints constituting the ideal conditions of choice *themselves* are constructed because they are committed to the idea that no normative truth exists independently of the practical

¹² Cf. Shafer-Landau 2003, 41–3.

standpoint. If they say that these normative constraints are *constructed* under appropriately specified conditions of choice, then they must further specify whether these conditions involve further normative constraints, and if so, how are they justified.

Idealized stance constructivists cannot avoid circularity or an infinite regress unless they help themselves to stance-independent, unconstructed normative facts.¹³ This, of course, would make them realists rather than constructivists.¹⁴

Can the procedural constructivist explain the metaethical status of the normative input by appeal to Railton's (1986) or Smith's (1994) naturalistic moral theories that define moral truth in terms of desires or psychological responses of a fully rational and informed observer under certain hypothetical conditions? Such ideal observer theories too take moral truths to be constructed from an idealized stance, after all. The only difference seems to be that they define what is good or morally right in terms of desires or responses of fully informed and coherent agents rather than placing emphasis on a procedure of construction. Both Railton and Smith adopt a reductive naturalistic approach to normativity and identify normative (or moral) value with what is *non-morally* good for the agent.¹⁵ In their view, the normativity of agents' non-moral good must be reduced to natural facts about what agents would desire themselves to desire if they had "complete and vivid knowledge of [themselves] and [their] environment" (Railton 1986, 174). Ideal observer theories are motivated by the rationalist idea that "the existence of reasons

¹³ Cf. Timmons 2003, 401; Raz 2003, 358.

¹⁴ Such a view would still be constructivist in normative ethics despite being realist in metaethics.

¹⁵ Railton 1986, 173–5; Smith 1994, 202.

presupposes that under conditions of full rationality we would all have the same desires about what we are to do in the various circumstances we might face.”¹⁶

The problem is that there seems to be no real *metaethical* difference between procedural constructivism and ideal observer theories, and thus the latter is similarly subject to Shafer-Landau’s Euthyphro-style objection. How is reducing the normativity of an agent’s non-moral good to natural facts about what agents would desire themselves to desire if they were fully informed and coherent different than reducing the normativity of a fair system of cooperation to natural facts about the decisions made under the ideal conditions of the original position?¹⁷ There seems to be no difference for our purposes because just as the procedural constructivist must account for the normativity of the input of the constructive process, so the ideal observer theorist must account for the normativity of the “conditions of full rationality.” Unless such an account is given in a non-question-begging way, *none* of the positions that associate normative or moral truth with *idealized* conditions can qualify as a distinct metaethical view. And prospects for such an account look pretty dim indeed.

The Euthyphro dilemma, then, poses a problem for *any* constructivist view that appeals to idealized conditions for creating normative truth. Procedural constructivists like Rawls and Scanlon and ideal observer theorists like Railton and Smith both emphasize the importance of an *ideal* deliberative process in determining normative facts. They all think normative truth is to be generated under idealized circumstances of practical reasoning, yet they characterize such circumstances differently. No matter how

¹⁶ Smith 1994, 198.

¹⁷ Cf. Street 2010, 372.

the ideal conditions of reasoning are characterized, though, the idealized stance constructivist *cannot* claim that *all* normative facts are constructed. If a deliberative procedure or process is to create *all* normative facts, then *no* consideration can have a normative weight independently of the procedure. However, we must regard some considerations as giving us good reasons to act or think in certain ways rather than in others in order to deliberate at all. If a cognitive process is not guided by *some reason*, then it merely involves arbitrary changes in mental states, and we cannot call such a process deliberation. Therefore, if the idealized stance constructivist claims that *all* reasons follow from a deliberative process, then deliberation becomes impossible because reasons are necessary for deliberation.¹⁸ Such constructivists thus need considerations that do not depend on a prior answer to the question of what the procedure generates to escape the objection. This is the reason idealized stance constructivists are also ‘local’ or ‘restricted’ constructivists, i.e., they make substantive *normative* judgments and regard them as *unconstructed* inputs of whatever constructive procedure they put forward. This concession, however, comes at the cost of not being able to present a distinct metaethical view.

3.2.4 Conclusion

The above considerations suggest that theories trying to find reasonable grounds for agreement by appealing to our self-conception as finite, rational, free, mutually

¹⁸ Cf. Enoch 2009, 332–3. This is also called the *bootstrapping* objection. The idea is that if no reason exists prior to a deliberative process then you are left with no resources but the process itself in explaining what reasons we have. However, you must *assume* certain reasons to be able to talk about a deliberative process in the first place. And since you have no other resources, you are pulling reasons out of nothing.

interdependent, and equal beings, or, for that matter, any position that appeals to idealized conditions in creating moral or normative truth must either have a *realist* (stance-independent) conception of objectivity, or they must weaken the standard of objectivity by focusing on intersubjective agreements. So, they either give up being *metaethical* constructivists, or they give us a weak, coherentist form of objectivity.¹⁹

3.3 Constitutivism

As we have seen, idealized stance constructivism asserts that normative or moral truth is generated out of an idealized decision procedure or the stance of an ideal agent. On such views, the procedure of construction or the ideal conditions of choice involve(s) substantive (unconstructed) normative judgments or moral values, which subjects the view to the Euthyphro-style dilemma regarding the order of determination.

Constitutivism avoids this dilemma because it characterizes the standpoint constituting the standard of correctness for our moral judgments as involving *no* substantive normative judgments or moral values. Rather, according to constitutivism, moral truth is determined by the *constitutive* features of agency/practical reason/valuing.²⁰ In particular, normative or moral content is extracted from the constitutive features of a *formally characterized* stance. Giving a stance a formal characterization amounts to claiming that a moral statement is true if and only if it is logically or instrumentally entailed by what is

¹⁹ Rawls adopts such coherentism. See Rawls 1999, 524; 2000, 268–73.

²⁰ A feature is constitutive of agency/practical reason/valuing, if *every* person that acts, or engages in practical reasoning or valuing, possesses it *merely in virtue of* being an agent/practical reasoner/valuer.

involved in acting/practical reasoning/valuing *as such*.²¹ In other words, constitutivism gives an account of what is involved in rationality as such or in acting or valuing anything at all. Since this explanation does not presuppose any particular normative judgment or moral value, constitutivism can be described as a distinct metaethical view unlike idealized stance constructivism.²²

Proponents of this type of constructivism argue that action, practical reasoning, or the attitude of valuing has a constitutive aim or principle that provides reasons for agents' actions even before they start making decisions. Since the proponents of this type of constructivism hold that moral content is *not* created by any procedure that involves decisions or by the stance of an ideal agent that involves desires but are conceptually derived from the constitutive features of agency/practical reason/valuing, the term 'constitutivism' has been used to distinguish this position from other types of constructivism. Constitutivism, then, is a distinct form of constructivism. Typically, each constitutivist theory makes three claims: (1) Constitutive Claim: *description* of the constitutive features of agency/practical reason/valuing, (2) Normative Claim: an account of *why* the nature of agency/practical reason/valuing has a normative significance, and (3) Content Claim: a list of first-order normative conclusions that follow from the nature of agency/practical reason/valuing.²³ Differences among constitutivist theories are due to their different takes on each claim.

²¹ Cf. Street 2012, 40.

²² Cf. Street 2010, 367.

²³ Cf. Bukoski 2016, 117.

3.3.1 Velleman

What does it mean to say that moral truth is determined by the constitutive features of action/practical reasoning/valuing? Let's look at some examples. Velleman (2009) argues that reasons for actions are entailed by the constitutive features of agency. On his account, what is constitutive of agency is the aim of self-understanding. His Constitutive Claim can be seen as an explanation for a certain conception of intentional action. According to the conception of intentional action Velleman (2009, 129–133) espouses, intentional action conceptually involves the agent's *immediate* knowledge of what she is doing. By "immediate" knowledge, Velleman means that the agent does not need to rely on evidence or introspection to know what she is doing.²⁴ For example, when I intentionally sing my favorite song at a karaoke party, I don't need to analyze the movements of my mouth, jaw, tongue, and lips, nor I need to reflect on my feelings associated with me singing that song in order to know what I am doing. I simply non-observationally know that I am singing my favorite song at a karaoke party.

Velleman proposes a theory of action that he thinks entails and best explains this "immediate, non-observational knowledge" conception of intentional action. According to Velleman's theory of action, *all* agents, solely in virtue of being agents, aim at self-understanding, and this is true independently of agents' starting set of normative judgments or values. That is, it is a conceptual truth that action involves the aim of self-understanding and, as the constitutivist idea goes, the standards of correctness for moral judgments are determined by factors under which agents' actions would make folk-

²⁴ Cf. Katsafanas 2018, 370.

psychological sense to them. Or, to put it differently, reasons for actions are simply considerations in light of which an agent could understand what she is doing.²⁵ Suppose an agent has a desire to understand what she is doing. Without a doubt, she will be inclined to act in ways that would make sense to her, and she will refrain from acting in ways that would not. For example, I know that I want to use my laptop to check the news right now. If I turned on my laptop and checked the news, my action would make sense to me. But if I threw my laptop against the wall and broke it into pieces, I would not understand what I was doing. So, according to Velleman's theory of action, I am more disposed to turn on my laptop and check the news than break it into pieces, precisely because of my desire to understand myself acting.

Velleman's theory of action might strike one as counterintuitive because we tend to think that we fulfill our desire to understand or know our behavior first by observing our behavior and then by forming beliefs in accordance with what we observed.²⁶ But, on Velleman's account, this is not the only way we can fulfill our desire for self-understanding: we *often* fulfill this desire first by forming beliefs and *then* by adjusting our behavior in accordance with these beliefs.²⁷ In particular, our desire to understand what we are doing prompts us to form *expectations* about our upcoming actions. And once we form such expectations, our desire to understand our actions will also prompt us to adjust our actions in a way to meet those expectations. This means that our

²⁵ "Considerations weigh in favor of an action, I propose, insofar as they contribute to an overall understanding of the action, given how the agent conceives of himself and his situation" (Velleman 2009, 19).

²⁶ Cf. Katsafanas 2013, 70.

²⁷ Velleman 2006, 224–52.

expectations about our upcoming actions are *self-fulfilling*: we are able to perform a certain action merely in virtue of forming an expectation that we will perform that very action.²⁸ For example, if I form the expectation that I will use an umbrella whenever it rains, I will use an umbrella the next time it rains to fulfill my desire for self-understanding. Or, I might have conflicting desires: some of my desires could tell me to go to the beach to relax, while others could tell me to stay home and watch a movie. Although I can fulfill my desire for self-understanding by performing either action, my desire to know what I am doing will prompt me to perform the latter action if I form the expectation that I will stay in and watch a movie.

Velleman believes that the “immediate, non-observational knowledge” conception of intentional action is best explained by characterizing intentional action as caused by self-fulfilling beliefs or expectations. That is, the immediate, non-observational knowledge we have about our intentional actions stems from self-fulfilling expectations we form about our upcoming actions: “the agent attains contemporaneous knowledge of his actions by attaining anticipatory knowledge of them” (Velleman 2004, 277). Velleman’s hypothesis also explains why an intentional action *always* entails immediate/non-observational kind of knowledge: if intentional actions *are* nothing but behaviors that are grounded in self-fulfilling expectations/beliefs, then we will *always* have immediate knowledge about our intentional actions. According to Velleman, these considerations about action support the idea that action has a constitutive aim of understanding what one is doing.

²⁸ Velleman 2006, 213–6.

Velleman's Constitutive Claim is that the aim of self-understanding is a constitutive feature of agency that provides reasons for actions. His Normative Claim is that the aim of self-understanding justifies itself. Velleman asserts that reflection on whether the aim of self-understanding is normatively significant is structured by the that very aim. In other words, the aim of self-understanding must be in effect in order for us to reflect in the first place.²⁹ To question whether the constitutive aim is justified "demands that the constitutive aim of action be justified in relation to the criterion set by the aim itself" (Velleman 2009, 138). If the aim of self-understanding is constitutive of agency, then, Velleman claims, it makes folk-psychological sense for us to consider that aim to be normatively significant or justified.³⁰ Velleman's Content Claim is that moral content is determined by whatever promotes understanding of what agents are doing. Velleman characterizes the aim of self-understanding as being "merely pro-moral," meaning that it does *not* always provide "moral" reasons for actions: "[t]here was no antecedent guarantee that [a moral] way of life would develop among rational agents, much less that moral conduct will be rationally required of every agent at all times" (Velleman 2009, 2). That is to say, immoral actions could make more folk-psychological sense for some agents, depending on their character and situation.

²⁹ Velleman 2004, 292–93.

³⁰ Cf. Bukoski 2017, 2672.

3.3.2 Street

Not all constitutivists base their theories on what is constitutive of agency. Street describes her version of constitutivism with a focus on the constitutive features of the attitude of valuing rather than action itself. Street defines the attitude of valuing as “taking oneself to have a reason” (Street 2008, 228). Taking oneself to have a reason has *three* constitutive features. First, taking oneself to have a reason to do *A conceptually* involves taking oneself to have a reason to take what one acknowledges to be the necessary means to *A*. Just as a bachelor who is married is *not* a bachelor, so someone who is judging that she has a reason to do *A* and also that she has *no* reason to take what she acknowledges to be the necessary means to *A* is *not* taking herself to have a reason to *A* (ibid.). Second, the attitude of valuing involves much more emotional and phenomenological complexity than the attitude of mere desiring. While desiring is strongly associated with pleasure or sensual gratification, valuing could involve a wide array of emotions and feelings such as displeasure, anxiety, determination, agony, courage, and so on.³¹ And third, valuing has a “greater structural complexity” and “much more complex attitudes towards the world” than desiring. Unlike desiring, which is usually directed at a particular object or a state of affairs, valuing involves “experiencing very specific features of the world as ‘calling for’ or ‘demanding’ or ‘counting in favor of’ other very specific things” (Street 2012, 44). That is, there is a difference between simply desiring an ice cream and experiencing a person telling a lie as calling for being skeptical about the sincerity of that person’s statements in the future.

³¹ Street 2012, 44.

After making her Constitutive Claim, Street moves on to explaining the justification of particular values. She concurs with the Kantian idea that “[i]f you value something, then you cannot – simultaneously, in full, conscious awareness – also think that there is *no reason whatsoever* to value it” (ibid., 46).³² In other words, any particular object or end we value requires a justification, i.e., we must explain the reason for valuing that object or end. However, she disagrees with Korsgaard’s regress argument,³³ according to which the requirement of justification leads to a reflective regress that comes to an inevitable end when we ask whether to value humanity, due to the literal practical necessity of valuing humanity for acting at all. Instead, Street argues that particular values one adopts are justified by her own further values.³⁴ This means that there is no independent standard of correctness for our moral claims once we step back from the whole set of our values. Rather, the only standard or principle that is entailed by the constitutive features of valuing is that of *coherence*. That is, when we try to justify particular values we adopt, we seek no more than to reach a “coherent web of interlocking values,” which signifies the end of the reflective regress.³⁵ Moral content is, then, whatever that constitutes a “coherent web of interlocking values” for an agent.

³² “We need reasons because our impulses must be able to withstand reflective scrutiny. [...] ‘Reason’ means reflective success” (Korsgaard 1996, 93, 97).

³³ Korsgaard 1996, 120–126; 2009, 20–25.

³⁴ Street 2008, 220–223; 2012, 51–52.

³⁵ Street 2012, 51.

3.3.3 The Basic Constitutivist Strategy and Inescapability

Constitutivism is ambitious in that it attempts to ground normative or moral truths in a less controversial descriptive foundation. We can see this strategy both in Velleman and Street. While Velleman gives an account of what is constitutively involved in acting intentionally, Street talks about the constitutive features of valuing or taking something to be a reason. Without a doubt, their accounts (their Constitutive Claim) can be contested; however, such an approach is different from basing objectivity on non-natural entities. They both try to make minimal assumptions about the constitutive features of the activity in question (e.g., self-understanding or the instrumental principle) and to derive normative pressures from those features. As long as one has a reason to (or must) engage in action or valuing, one should adhere to its constitutive standards. Otherwise, one would lose the activity in question, and this is a price one would not want to pay.

This strategy is trivial and powerful at the same time.³⁶ It is trivial because if you engage in an activity, it is obvious that you must engage in it as it is rather than something else. This is just what the concept of identity entails. For example, if I have a reason to play baseball, I must comply with the rules of baseball. If I don't, then I am not playing baseball anymore. Of course, if I don't have a reason to play baseball to begin with, then I don't need to worry about the rules of baseball. The skeptic may press this point and say that the basic constitutivist strategy has a *conditional* character: if we don't have a reason to act or value, then we don't need to worry about the constitutive standards that govern such activities. In other words, providing the metaphysical foundation of an activity A, i.e., showing what is constitutive of A, does not give us the

³⁶ Cf. Ferrero 2019, 2.

source of authority. Constitutive standards carry only *formal* normativity. To show that these standards also carry *substantive* normative authority, the constitutivist must do more than merely transmit the pressure to engage in A to the constitutive standards of A. If we *already* have a reason to engage in A, then the basic constitutivist strategy could be useful as a transmission device. However, it seems that the constitutivist needs something beyond this basic strategy to locate the source of authority within A.

The basic constitutivist strategy seems also to be powerful. First, it is not fully trivial because it takes some philosophical effort to provide the metaphysical basis of the constitutive standards of an activity, as the examples of Velleman and Street show. Second, constitutivists do not talk about constitutive standards of games or etiquette. Rather, they talk about action or the attitude of valuing, and even the skeptic seems to be in trouble if she ever loses these activities. In fact, even when the skeptic raises her worry, she is already engaged in valuing and acting. It could be that the activities that constitutivism describes are *inescapable*. Korsgaard might be correct in saying that “[h]uman beings are *condemned* to choice and action” (2009, 1). If valuing and acting are inescapable aspects of the human condition, then perhaps we should all worry about losing these activities. Otherwise, we would lose ourselves. The idea is that if we all (should) have this worry simply by virtue of being human or rational, then the constitutivist could locate the source of authority within action or valuing. And the hope is that the question, “Why be agents?” becomes unintelligible because we have no option but to be agents.

The self-application of the basic constitutivist strategy and the appeal to inescapability, however, may not be powerful enough to defuse the conditionality

objection by themselves. First, even though the activity of engaging with ourselves could be necessary to act or value anything at all, the source of the authority of self-engagement could be more specific things that we value or simply biological fitness.³⁷ Second, it could be that the reason to mind one's self-loss is independent of the constitutivist strategy itself. Even if we assume that there is a categorical reason to value our own existence, this reason does not appear to arise from the self-application of the basic constitutivist strategy itself. For example, it is possible to ground the reason to mind one's self-loss in an absolute inner value that a realist would be happy to accept. Third, even if the question "Why be agents?" can only be asked from within agency, it still may be possible for the skeptic to take up a stance outside of her engagement with agency when she asks that question. In other words, it may be dialectically possible for an agent to question the normative force of being an agent. This is Enoch's (2006) objection to the self-application of the basic constitutivist strategy and to the appeal to inescapability. If constitutivism cannot overcome this conditionality or dependency problem, then it is possible to ground the authority of being an agent or valuing in non-natural entities that realism postulates. Thus, to be able to claim that constitutivism is a plausible alternative to realism, one must meet the dependency challenge within the constitutivist framework. Let's explore two main versions of constitutivism to see if they can meet the challenge.

3.3.4 Humean vs. Kantian Constitutivism

Although both Humean and Kantian versions of constitutivism agree that normative/moral content is derived from the constitutive features of agency/practical

³⁷ Cf. Ferrero 2019, 8.

reason/valuing, they disagree on *how* that content is derived. Suppose a person who values above all else molesting children. Suppose also that his valuing molesting children is consistent with all of his other values and all non-normative facts. What should we think of this ideally coherent agent? Should we say that he fails to recognize *stance-independent* reasons that tell against committing such an abominable act, or should we simply reject the possibility of such an agent? The moral realist makes the former claim, and this is what distinguishes his position from constitutivist ones, which reject that claim (both Humean and Kantian kinds). What distinguishes Humean constitutivism³⁸ from Kantian constitutivism³⁹ is that while Humean constitutivists believe that an ideally coherent immoral agent can exist (recall Street's aim of reaching a "coherent web of interlocking values"), Kantian constitutivists reject that possibility. In other words, moral failures are *rational* failures, according to the Kantian, because substantive moral reasons follow from the nature or structure of agency/practical reason/valuing *as such*.

Humean constitutivists find the Kantian picture unsatisfying. The question is, how can a Kantian vindicate anything substantive if she resorts to a minimal (and formal) conception of agency or practical reason? It is precisely this bootstrapping (or emptiness) objection that drives Humean constitutivists to reject the idea that the structure of agency/practical reason/valuing, devoid of any particular evaluative standpoint, entails moral reasons. Rather, they argue, moral reasons are a function of contingent evaluative starting points. As Street claims, we find no independent standard of correctness for our

³⁸ Street's (2008, 2010, 2012) and Velleman's (2009) constitutivist accounts, among others, are considered to be Humean.

³⁹ Korsgaard's (1996, 2009) and Sensen's (2013, 2017) constitutivist accounts, among others, are considered to be Kantian.

moral judgments once we step back from the whole set of our values. Similarities in our reasons may be a result of similarities in our initial set of normative judgments and conditions or a shared human nature. So, if we had different circumstances or different nature, our reasons would change accordingly.

Humean constitutivism differs from standard Humean views in that it does not identify constitutive features of agency with *desires* but with *normative judgments*.⁴⁰ The reason is twofold. First, if Humean constitutivism attempted to derive normative judgments from a non-normative input such as desires, then it would not be able to present itself as a distinct metaethical option, just like idealized stance constructivism. This is because one needs a substantive (rather than formal) normative claim to connect a non-normative input and a normative output, as we have seen in 3.2.2. Humean constitutivism is a *non-restricted* (as opposed to restricted or local) kind of constructivism, which means that it applies to *all* normative judgments. In other words, the correctness of *all* normative judgments is constructed from “a standpoint constituted by some further set of normative judgments” (Street 2008, 220). Second, Humean constitutivism attempts to derive normative reasons from the agent’s perspective, and appealing to (first or higher order) desires or an agent’s subjective motivational set like Williams⁴¹ makes it difficult to distinguish the agent’s standpoint from fleeting motivational forces operating within an agent. An agent could be alienated from some of the elements in her motivational set, and

⁴⁰ Cf. Street 2008, 245.

⁴¹ An agent’s subjective motivational set involves “dispositions of evaluation, patterns of emotional reaction, personal loyalties, and various projects [...] embodying commitments of the agent” (Williams 1981, 105).

some of the elements could represent her standpoint more deeply than other elements.⁴² Due to the need for a more fine-grained understanding of the constitution of the agent's standpoint, Humean constitutivism regards normative truths as being constitutively entailed by the agent's *actual normative judgments* in combination with the non-normative facts.⁴³

Humean constitutivism is moderately Kantian: it entertains the notion of giving laws to oneself (or autonomy) in understanding moral truth. In that respect it bears similarities to the 'weak externalist' or 'responsiveness-to-reasoning' accounts of personal autonomy.⁴⁴ According to the weak externalist conception of self-governance, distancing oneself from one's current motives and choosing between them or endorsing them is necessary but not sufficient for self-governance or self-determination. One must also evaluate one's motives based on one's further beliefs and desires and adjust these motives in accordance with one's evaluations to govern herself. That is, on this view, there is more to autonomy than the capacity to hold higher-order attitudes (or to endorse one's motives): the capacity to discern what is entailed by one's beliefs and desires is also crucial to autonomy.⁴⁵ Admittedly, Humean constitutivism (Street's version) begins with normative judgments as the inputs to constructivist scrutiny instead of beliefs or desires. Nevertheless, the Humean constitutivist's focus on the agent's ability to draw inferences from her initial standpoint and her ability to reconsider her motives in

⁴² Cf. Bratman 2012, 85.

⁴³ Street 2008, 232.

⁴⁴ Examples of this view are Christman (1991, 1993) and Mele (1993, 1995).

⁴⁵ Cf. Buss and Westlund 2018.

accordance with what follows from her initial standpoint captures the weak externalist conception of personal autonomy and the *self-legislation* aspect of Kant's autonomy.⁴⁶

Humean constitutivism is only moderately Kantian because it does not capture the *causal independence* aspect of Kant's autonomy. In Kant's view, moral reasons are *categorical* reasons, which are created autonomously. Kant's conception of autonomy indicates a capacity to behave independently of the necessitation of natural causes, especially of desires and inclinations.⁴⁷ A reason is categorical when "I ought to act in such or such a way even though I have not willed anything else" (GMS 4:441). So, acting in accordance with a 'categorical ought' requires that one is not governed by natural laws but is governed by the law of her reason, namely the moral law. According to Kant, one is autonomous only when one's will determines itself by the commands of *pure* reason, which are abstracted from everything empirical. This is why the mere self-legislation does not show why autonomy is the source of (or the supreme principle of) morality on Kant's view. Since Kant identifies autonomy with *pure* reason, which is "cleansed of everything empirical" (GMS 4:388f.), Kant's conception of autonomy involves both the notion of self-legislation and the notion of causal independence.

The difference between Humean and Kantian versions of constitutivism could be spelled out in terms of their different views about the way in which normative/moral reasons are generated by the constitutive features of agency/practical reason/valuing. Humean constitutivists believe that *aims* are sufficient for producing moral reasons. If we inescapably aim at Ψ -ing whenever we act, or if action conceptually involves aiming at

⁴⁶ Cf. KpV 5:31f.; KrV B1f.

⁴⁷ Cf. KrV B479, B561; GMS 4:389, 4:440; KpV 5:33.

Ψ -ing, then Ψ -ing constitutes a standard of success for action and generates reasons for it, on the Humean model. For example, as we have seen above, on Velleman's view, the aim of self-understanding constitutes a standard of success for action, and on Street's view, the aim of reaching a "coherent web of interlocking values" generates reasons for actions. Kantians disagree. They believe that aims, and desires connected with them, are *not* sufficient for providing moral reasons because they are external to the will and grounding moral normativity in them would bring about *heteronomy*.⁴⁸ For that reason, Kantian constitutivists focus on constitutive *principles* rather than constitutive *aims*. For example, Korsgaard (1996, 2009) attempts to show that Kant's Categorical Imperative is the constitutive principle of action, and then she derives moral content from that principle, without appealing to a constitutive aim. We can, therefore, say that Humeans support "aim-based" versions of constitutivism, whereas Kantians support "principle-based" versions of it.⁴⁹

3.3.5 Problems with Humean Constitutivism

According to Humean constitutivism, normative or moral reasons follow from an agent's standpoint in combination with the non-normative facts. It is, however, not clear what

⁴⁸ As mentioned above, a reason is categorical when "I ought to act in such or such a way even though I have not willed anything else" (GMS 4:441). If we derived moral reasons from aims, be they constitutive of action or not, they would apply to us only hypothetically and contingently: "I ought to do something *because I will something else*" (ibid.). "The will in that case [would] not give itself the law; instead the object, by means of its relation to the will [would give] the law to it" (ibid.). And "[i]f the will seeks the law that is to determine it *anywhere else* than in the fitness of its maxims for *its own giving* of universal law [...] *heteronomy* always results" (ibid.; emphasis added).

⁴⁹ Cf. Katsafanas 2018, 377.

constitutes an agent's standpoint. For example, Bratman (2012), another Humean constitutivist, criticizes Street's identification of the agent's standpoint with normative judgments. First, he claims that we can be alienated from some of our normative judgments, just as we can be alienated from some of our desires.⁵⁰ He gives the example of Huckleberry Finn's judgment that he should turn in Jim, the runaway slave. Huck Finn is alienated from that judgment due to his attitudes in support of protecting him. This means that not all normative judgments constitute the agent's standpoint. Second, Bratman asserts that certain attitudes or conative commitments such as caring and love play an important role in shaping one's standpoint: "[t]he role of my love for my children in constituting my standpoint is not exhausted by judgments about reasons that are not themselves grounded in my love" (Bratman 2012, 92). For instance, when one makes a career decision after reflecting on options that seem equally interesting, she has a commitment that constitutes her standpoint, but that commitment goes beyond her normative judgments that do not depend on that commitment. The problem is that, as discussed above, if we begin with conative commitments as the inputs to constructivist scrutiny, then we need substantive normative judgments to derive normative (or moral) reasons. But then the view becomes normative rather than metaethical.

Can Humean constitutivism give us a satisfying account of moral objectivity? On the Humean constitutivist view, morality is objective in the sense that the constitutive standards of agency or the attitude of valuing is universal. For example, Velleman argues that *all* agents, *qua* valuers, aim at *self-understanding*, and this is true independently of their starting set of normative judgments. According to Street's Humean constitutivism,

⁵⁰ Bratman 2012, 91–2.

constitutive features of agency entail that *coherence* is the only standard of correctness for our moral judgments. The fact that Humean constitutivism espouses a *coherentist* account of moral truth constitutes a problem for our purposes. This is because, even though the standards of correctness for our moral judgments are laid out by what is entailed from the constitutive features of agency or valuing, moral truths are essentially a function of one's contingent normative starting points. Such a view does not provide us with a satisfying account of moral objectivity that allows us to call a coherent psychopath immoral. Since Humean constitutivism can only give us a *weak* sense of objectivity that associates moral truth with *contingent* initial set of values, it either collapses into *relativism* about moral truth (a minor change in the initial set of values could in principle create a unique, coherent moral reality for *each* individual or society) or it has to make a highly ambitious and probably incorrect empirical claim (that each individual's or society's initial set of values are identical to each other).

The claim that Humean constitutivism collapses into relativism appears to be inconsistent with my previous claim that it is a distinct metaethical position. However, my claim is not that Humean constitutivism is a fully relativistic theory. After all, Humean constitutivists argue that each agent, solely in virtue of their capacity to make normative judgments, inescapably has a constitutive aim that provides reasons for actions. On the one hand, this view is nonrelativistic because it allows for *errors* in the content of one's moral judgments: if a particular moral judgment one makes is inconsistent with her other moral judgments, or if it is inconsistent with the constitutive aim, then it is *wrong*, according to the Humean constitutivist. This aspect of Humean constitutivism distinguishes it from subjectivism and relativism, according to which there

is no standard of correctness for our moral judgments. On the other hand, the position is still relativistic because the output of practical deliberation varies in accordance with the agent's initial standpoint, regardless of whether it is constituted only by normative judgments or by normative judgments *and* conative commitments. This indicates the possibility of many and conflicting moral truths. Humean constitutivism has an unappealing implication that moral judgments of serial killers or people who support genocide or female genital mutilation reflect moral truths provided that their judgments constitute a coherent set and/or are in line with the constitutive aim of agency.

Humean constitutivism does not have enough resources to close the explanatory gap between the Constitutive Claim and Normative Claim of constitutivism due to the sort of objectivity they support. The question is, *why* should we regard merely descriptive features of agency as normatively significant? Or, *why* shouldn't we regard our endorsement of the constitutive aim as having solely an *instrumental* value? Even if we concede that constitutive features of agency determine the standards of correctness for our moral judgments, we do not have strong reasons to believe that these standards have a non-derivative, normative value. The constitutive aim could justify itself given that the agent does not have any conflicting normative commitments. There is at least no obvious reason to reject the normative significance of the constitutive aim under the circumstances where one endorses *having* the constitutive aim. It makes sense to endorse having the constitutive aim because this will make it more likely to fulfill that aim. Nevertheless, according to Humean constitutivism, what the constitutive aim prescribes depends to a great extent on one's normative or evaluative starting points. Thus, the constitutive aim can only be justified in an instrumental manner.

For instance, think of a psychopath and a hedonist. Suppose that the psychopath has the initial normative judgment that it is right to make people suffer, and that the hedonist has the initial normative judgment that maximizing her pleasure and avoiding pain are the only components of her well-being. The central project of the psychopath is to make people suffer, and the hedonist's project is to maximize her pleasure and minimize her pain. Now, it makes sense to both to endorse *having* the constitutive aim, be it self-understanding or coherence or some other aim, because (1) they must have that aim to be an agent at all, and (2) they must be an agent for their projects to succeed. However, it seems that endorsing the constitutive aim does not have a non-derivative value. Rather, its value seems to be derived from its contribution to their respective projects.

For example, suppose that the psychopath makes two further judgments: (PJ₁) "I am cognitively superior to other people," and (PJ₂) "Other people deserve to get tortured or killed by me." Suppose also that the hedonist makes two further judgments: (HJ₁) "Torturing an innocent person is morally right given that the torturer's pleasure surpasses the victim's pain," and (HJ₂) "Cheating on one's partner is morally right given that the cheater's pleasure surpasses the partner's pain." Making judgments PJ₁ and PJ₂ makes it easier for the psychopath's actions to make folk-psychological sense to him, and also, PJ₁ and PJ₂ contribute to the coherence of the starting set of his evaluative judgments. Moreover, making judgments PJ₁ and PJ₂ facilitates achieving his aim of making people suffer because both judgments contribute to the justification of that aim and thereby motivate the execution phase of his central project. The same considerations are applicable to the tripartite relationship between HJ₁ and HJ₂, the constitutive aim of agency, and the aim of the hedonist's project. Thus, endorsing the constitutive aim (of

self-understanding or coherence) helps both achieve the aims of their general projects. However, it is not clear why the constitutive aim of agency would have non-instrumental, *moral* significance.

Furthermore, Humean constitutivism does not seem to allow for the possibility of self-transformation and genuine autonomy. Can an ideally coherent serial killer change his mind upon discovery of a good reason, such as that killing people is wrong and that becoming a law-abiding citizen is right? It is not very difficult to imagine such a possibility. This would perhaps be an extreme example, but life is full of less dramatic examples of this sort. People can change their moral outlook after reading an interesting book, having an interesting conversation, or having an enlightening experience, and they at least sometimes change their actions accordingly. But such moral changes do not produce genuine reasons for action on this strictly coherentist Humean view because they threaten the diachronic stability and consistency among one's judgments. In fact, the agent's initial standpoint seems to have a despotic status, just like an authoritarian constitution that cannot be changed. As Rousseau says in *The Social Contract*, "if the established order is bad, why should the laws which prevent its being good be regarded as fundamental" (Rousseau 1968, 99)?⁵¹

My claim is not the dogmatic conclusion that Humean constitutivism is wrong. Humean constitutivism may not meet the conditionality or dependency challenge; however, it may still show how our agency, or the attitude of valuing, is constitutively connected to other important aspects of our existence. For example, on Bratman's (2018) view, agency is constitutively connected to our sociality and self-governance, and it is

⁵¹ Cf. Shemmer 2012, 176.

inescapable because of its deep entrenchment in much of what we take to be important to us. In Walden's (2018) view, agency is constitutively connected to our embodiment and socialization, which are inescapable despite being metaphysically contingent. Such Humean accounts can inform us about the structure of agency and how entangled normative reasons are with reasons derived from other activities and capacities that are (allegedly) constitutively connected to our agency. They, however, cannot provide us with a satisfying account of moral objectivity.

My claim is that people who are dissatisfied by the relativistic tendencies, instrumental rationality, and strict coherentism of Humean constitutivism do not have to adopt the problematic realist picture to ensure moral objectivity, since Kantian constitutivism, which is bolstered by Kant's autonomy-based objectivity, could close the gap between the Constitutive Claim and Normative Claim without having to resort to a stance-independent moral reality.

3.4 Kantian Constitutivism

In this section, I claim that Kantian constitutivism can give us a strong account of objectivity unlike Humean constitutivism, but that it does so without appealing to stance-independent moral properties or facts like realism. I discuss two main versions of this type of constitutivism: (1) Christine Korsgaard's (2009) project of associating morality with agency-enabling principles that provide psychological unity, and (2) Oliver Sensen's interpretation of Kant's metaethics as transcendental constitutivism, according to which moral value is not a stance-independent property that is 'out there' or tied to our rationality as an absolute inner property but rather it is "what reason deems necessary" or

“how reason judges” (Sensen 2013, 67, 68). On the former view, the Categorical Imperative (henceforth, the CI) is constitutive of empirical willing, i.e., the psychological conditions for endorsing motives as reasons for action. On the latter view, the CI is constitutive of *pure* willing. That is, the CI governs our consciousness in thinking about practical matters. This means the CI *guides* rather than constitutes one’s choices. In this section, I explain why I prefer the latter version of Kantian constitutivism to the former one as an alternative to moral realism.

Both Korsgaard’s and Sensen’s versions of constitutivism take morality to be necessary and universal yet still dependent on human reason. On both views, principles of practical reason do not depend on contingent desires and attitudes; that is, moral requirements are *categorical* requirements of reason. Moreover, both views try to capture the objectivity and categorical normativity of morality by associating moral principles with a particular function of our capacity for reason: moral principles do not exist independently of human reason. So, both views try to establish that we are not alienated from moral truths in metaphysically and epistemically problematic ways. Thus, they are, *prima facie*, good candidates for the purpose of this dissertation, namely, to find a plausible alternative to moral realism.

They, however, focus on different functions of the faculty of reason when they say moral principles are constitutive of practical reason. While Korsgaard claims that the CI is constitutive of the *executive* function of reason, the CI is constitutive of the *legislative* function of reason on Sensen’s interpretation of Kant.⁵² Kant calls the executive function of reason *Willkür*, which refers to our capacity to choose between different motives and

⁵² See Allison 1990, 30 for the distinction between these two functions.

maxims (subjective principles of volition), and he calls the legislative function of reason *Wille*, which is the *source* of the moral law. That is, *Wille* generates the moral law and provides a standard for the selection of motives and maxims (MS 6:213–4). In Kant’s view, *Willkür* must be guided by *Wille*: what springs from the nature of *Wille* becomes an imperative for *Willkür*. *Wille* is what Kant calls ‘pure reason’ or ‘freedom.’

Pure reason [...] gives (to the human being) a universal law which we call the *moral law* (KpV 5:31f.).

What, then, can freedom of the will be other [...] the will’s property of being a law to itself? [...] This, however, is precisely the formula of the categorical imperative and is the principle of morality; hence a free will and a will under moral laws are one and the same (GMS 4:447).

To reiterate, Korsgaard identifies the moral law with *Willkür* or empirical willing, whereas Sensen’s Kant identifies the moral law with *Wille* or pure willing. This is the main difference between these two versions of Kantian constitutivism, and as I claim in this section, this is precisely why the latter is a neglected (and perhaps a better) alternative to moral realism.

I begin with a constitutivist (or anti-realist) reading of Kant’s moral theory (3.4.1), which I believe motivates both Korsgaard’s and Sensen’s versions. I then lay out Korsgaard’s constitutivism (3.4.2) and talk about the problems associated with it (3.4.3). First, I accept that Korsgaard’s appeal to inescapability successfully defuses Enoch’s skeptical challenge (or the dependency challenge). This is an important achievement, because it shows that the source of moral objectivity and categorical normativity does not lie outside of the capacity of practical reason. Despite this, as Ferrero (2019) notes, this is only a defensive move. It may be true that an *external* challenge is inconceivable, but the *internal* challenge remains: do we have compelling reasons to be agents? If we do, what is the nature of those reasons? In other words, the inescapability claim can only “sets the

boundaries for the explanation of the source of authority,” (Ferrero 2019, 14) but it does not explain how this authority takes place, which is precisely what the constitutivist must explain. It is perfectly possible that an absolute inner value property that is tied to our rationality generates this authority. But this would be realism.

Furthermore, Korsgaard’s claim is that the CI is constitutive of agency, but it is not as if it is *literally* practically necessary to follow the CI. Rather, we must at least *try* to do so (Korsgaard 2009, 45, 81). Korsgaard cannot resort to literal practical necessity when it comes to particular cases, since, otherwise, no immoral action would be possible. That is, when we act immorally, we do not disintegrate as agents, but our actions are *inconsistent* with the commitment we take on in acting at all.⁵³ But then the question is, what makes consistency or coherence morally relevant? Is consistency between our practical commitments and our actions *intrinsically valuable*? But that would be realism. Here again, Korsgaard cannot rule out realism. Worse yet, Korsgaard cannot establish that we must at least *try* to follow the CI in acting at all. Her arguments can at most establish that action constitutively involves trying to follow *subjectively* universal principles. For all one needs to do to function as an agent is to adopt principles that apply universally throughout *one’s own life*. For instance, a serial killer surely functions as an agent when he acts on a principle that gives him reasons to kill only human adult females. The combined effect of these factors prevents me from saying that Korsgaard’s constitutivism is a plausible alternative to moral realism.

As the problems with Korsgaard’s account show, we must leave room for immoral actions. It is perfectly possible for one to dismiss moral considerations and be a serial

⁵³ Cf. Fitzpatrick 2013, 57.

killer, a devious manipulator, a free-rider, and so on. One can still function as an agent when one does not act on morality's or reason's demands. However, one's actions cannot be called moral in those cases. While subjective principles constitute action, the CI constitutes *moral* action or how pure reason functions. In 3.4.4, I present a specific reading of Kant that supports this idea, a transcendental constitutivist reading. According to transcendental constitutivism, moral value or goodness is (a) not an absolute inner property that is tied to our rationality; (b) it is not the condition of the possibility of acting or choice (*Willkür*); or (c) it is not the condition of the possibility of *successful* rational agency in the sense of acting in accordance with one's practical commitments. Rather, moral value is "constitutive of how pure reason operates" (Sensen 2017, 218). In other words, moral value is conceived of as an operating principle of how our reason *functions* rather than as an *entity* or *property* that shows up on an ontological radar screen. This view captures the objectivity and categorical normativity of morality unlike other forms of constructivism and constitutivism, and it circumvents the problems associated with the ontological characterization of moral value (3.4.5) and the perceptual characterization of moral knowledge (3.4.6).

In 3.5, I respond to possible objections to transcendental constitutivism. First, transcendental constitutivism does not appeal to a noumenal realm to ensure objectivity and categorical normativity (3.5.1). Rauscher's (2015) distinction between independence from *causation* in nature and independence from *existence* in nature explains this idea very well: just as a wax cannot change the mold when they are in contact, so the content of the law prescribed by reason remains unaffected by natural causation, even though our reason resides in nature. That is, the fact that pure reason gives the moral law

spontaneously does not entail that pure reason is ontologically independent of the natural world. Rather, we can conceive of pure reason or freedom (*Wille*) as residing in nature as an emergent, unalterable structure of thinking. Admittedly, transcendental constitutivism is a non-trivial view about how reason functions. (Recall the basic constitutivist strategy to ground moral truths in a less controversial descriptive foundation.) And, as Kant also concedes (KpV 5:30), we have no conclusive proof that we are really free in the way he describes it. However, it is possible to find companions in guilt (or perhaps, the more precise label would be ‘companions in innocence’) that also place the *function* of reason at the center, such as Chomskyan linguistics, Mikhail’s (2007) universal moral grammar, functionalism in the philosophy of mind, and Fodor’s modularity of mind.

Second, the objection that no substantive moral content can be extracted from Kant’s CI is hasty (3.5.2). Admittedly, a purely formal law, considered in and of itself, is empty. However, if our reason draws inferences from *empirically identifiable* universal human ends to determine moral content and makes necessary means to those universal ends *binding* through the CI, the objection may fail. As Kant himself says, “moral laws are to hold for every rational being as such [... but one] needs anthropology for its *application* to human beings” (GMS 4:412).

Finally, as we have seen in the first and the second chapter, evolutionary debunking arguments (EDAs), *prima facie*, pose a serious epistemic threat to the non-naturalist picture of moral objectivity. However, transcendental constitutivism is compatible with the idea that our sense of moral objectivity and categorical normativity have promoted our survival and reproductive success (3.5.3). I defend the idea that our capacity to be spontaneous, or to be free from the influence of nature, namely pure reason or freedom, is

the source of morality, and that it could have been evolved in nature. However, our sense of moral ought, the intuition that we *can* and *should* choose the morally right thing despite our strongest desires, follows *necessarily* from freedom (*Wille*) itself rather than from the forces of natural selection. Having this moral sense could have had a positive effect on our biological fitness as Joyce (2001, 2006) says, but no evolutionary history is needed to account for the relation between pure reason and the CI, just as no evolutionary history is needed to account for the truth of “2+2=4.” This is because the CI arises *a priori* from reason (or freedom). That is, the CI is the *byproduct* rather than the direct product of evolution. If we could create a free being, a being with a mind that has a certain level of complexity, *out of nothing*, that being would be under the CI and would have a sense of moral ought. This picture is perfectly compatible with our evolutionary history.

3.4.1 Kant against Realism

Kant’s moral theory has both realist and constructivist interpretations.⁵⁴ Here I briefly talk about one way of reading Kant. In particular, I present Kant as someone who is against moral realism. My aim, however, is to assess the philosophical merits of Kantian constitutivism rather than proving that Kant himself supports such a view. It is therefore sufficient to show that there are indications for constitutivism in Kant’s moral theory.

Kant agrees with realism that in order for morality to be objective and to have an absolute authority over all of us, it must be strictly necessary and universal, i.e., it must

⁵⁴ See Rawls (1980), O’Neill (1989), Korsgaard (1996), and Sensen (2013, 2017) for constructivist readings of Kant, and see Kain (2004, 2017), Langton (2007), Wood (2008), Hills (2008), Bojanowski (2015), and Schönecker (2017) for realist readings.

be independent of contingent attitudes, and it must apply to all rational agents. According to Kant, morality “must carry with it absolute necessity; that, for example, the command ‘thou shalt not lie’ does not hold only for human beings, as if other rational beings did not have to heed it” (GMS 4:389).

Kant’s distinctive claim, however, is that grounding morality in *stance-independent* values cannot capture the necessity and universality of morality. He rejects both non-natural and natural properties as possible foundations for morality. First, Kant does not ground morality in a non-natural value property. Since our attempts to reach conclusions about *transcendent* realities, such as God and freedom, result in contradictions, we are not allowed to postulate non-natural entities or a special faculty of intuition to discover them:

We cannot originally cook up [...] a single object with any new and not empirically given property and ground a permissible hypothesis on it [...] Thus we are not allowed to think up any sort of new original forces, e.g., an understanding that is capable of intuiting its object without sense (KrV B798).

Moreover, “[i]f the concept of the good is not to be derived from an antecedent practical law but, instead, is to serve as its basis, it can be only the concept of something whose existence promises pleasure” (KpV 5:58). Since it’s not possible to determine *a priori* what causes pleasure or displeasure, moral value would then be up to one’s individual experiences. That is, value would then be “restricted to individual subjects and their receptivity” (ibid.). If that were the case, then “there would be nothing at all immediately good” (KpV 5:59). Rather, the good or value would only serve to achieve some sort of pleasurableness.

Second, Kant does not ground morality in a natural value, such as happiness, pleasure, or healthiness. A necessary and universal moral law cannot be based on pleasure because

what causes pleasure differs greatly from one person to another. That is, pleasure is relative and contingent: “From the feeling of a sensation that may be different in every creature, no generally valid law can be derived for all thinking beings” (VE 29:625). Happiness cannot ground morality either, for Kant thinks that happiness is an indeterminate concept that involves “a maximum of well-being in [one’s] present condition and in every future condition” (GMS 4:418). It is impossible to determine which actions are supposed to make one happy. Even if happiness were a determinate concept, it still could not ground morality because moral actions would not then be “objectively necessary of [themselves]” but would be relative to some further end, namely happiness (GMS 4:414). What makes people happy differ substantially among different people and also within oneself over time (KpV 5:25). Happiness thus can only be *subjectively* necessary.

On Kant’s view, stance-independent values (natural or non-natural) cannot ground necessary and universal moral standards because, otherwise, the reason for following those standards would depend on a *desire* to achieve something. That is, moral obligation would not be objective and categorically normative but merely a means to desire-satisfaction. Theories that base morality on *stance-independent* facts describe laws that should be followed because of “some interest by way of attraction or constraint, since it did not as a law arise from *his* will; in order to conform with the law, his will had instead to be constrained by *something else* to act in a certain way.” (GMS 4:433) Desires and feelings cannot ground a necessary and universal morality since they “by nature differ infinitely from one another in degree” (GMS 4:442f.).

What could then be the metaphysical ground of morality? According to Kant, if morality is to be universal and necessary, it should be grounded in *autonomy*. That means morality is a law of *one's own* reason:

By explicating the generally received concept of morality we showed only that an autonomy of the will unavoidably [...] lies at its basis (GMS 4:445).

Autonomy of the will is the sole principle of all moral laws and of duties in keeping with them; *heteronomy* of choice, on the other hand, not only does not ground any obligation at all but is instead opposed to the principle of obligation and to the morality of the will (KpV 5:33).

Therefore, the ground of obligation [...] must not be sought in the nature of the human being or in the circumstances of the world in which he is placed, but a priori simply in concepts of pure reason (GMS 4:389).

Kant's autonomy-based objectivity is an alternative – and perhaps a more plausible – way of accounting for a necessary and universal morality. As we have seen in the first chapter, postulating non-natural entities or properties as foundations for morality alienates us from moral truths in metaphysically and epistemically problematic ways. A less problematic way to account for necessity and universality could be to ground morality not in any entity or property but in the standpoint of reason, i.e., in how pure reason functions.⁵⁵

3.4.2 Korsgaard

Before I discuss the version of Kantian constitutivism I support, namely transcendental constitutivism, it is important to mention perhaps the most famous version of Kantian constitutivism and show why it does not work for our purposes. When someone utters the term 'Kantian constitutivism,' Korsgaard is probably the first person that comes to mind;

⁵⁵ I discuss the metaphysical basis of this view in 3.4.5.

however, when I say Kantian constitutivism is a neglected alternative to moral realism, I don't have Korsgaard's version of Kantian constitutivism in mind. I will explain why.

Korsgaard's main claim is that the CI is constitutive of action or our capacity for choice.⁵⁶ Her first step is to define constitutive standards and action: "Constitutive standards are ones that the object or activity must at least try to meet insofar as it is to be that object or activity at all" (Korsgaard 2009, 32). For instance, if we do not follow (or try to follow) the laws of logic, then we are not thinking. Or, if something is not a habitable shelter, then it is not a house. She then claims that laws of practical reason are constitutive of action: "laws of practical reason govern our actions because if we don't follow them, we aren't acting" (ibid.). That is, to function as an agent or to have psychological unity, one must endorse the legitimacy of some general principle that provides reasons for action: "goodness is not a goal for people, but rather is our name for the inner condition which enables a person to successfully perform her function – which is to maintain her integrity as a unified person, to be who she is" (2009, 35).

Being an agent consists "in the activity of constantly making yourself into a person" that chooses and endorses desires as reasons for action (ibid., 42). There is no real difference between the activity of a morally *good* agent and a morally *defective* or *bad* agent: both must follow the constitutive principles of being an agent. A morally bad agent engages in the same activity, i.e., he tries to conform to universal principles, but he does it badly (e.g., when he fails to take the means to the end he wills out of weakness of the will or fear). Korsgaard defines action as "a movement attributable to an agent considered as an integrated whole, not a movement attributable merely to a part of an agent, or to

⁵⁶ This is what Kant calls *Willkür*, namely the executive function of reason. See Allison 1990, 130.

some force working in her on her.” (ibid., 45) She says “to act is to constitute yourself as the cause of an end” (ibid., 72). But caution is needed here: if you fail to realize the end you are pursuing, you are still acting even though you do not constitute yourself as the cause of that end. Therefore, what Korsgaard means is that to act is to *try* to constitute yourself as the cause of the end you are aiming at.

Korsgaard’s second step is her argument against particularistic willing – the crucial aspect of her view (ibid., 72–80). Particularistic willing is “to have a reason that applies only to the case before you, and has no implications for any other case” (ibid., 72–3). It is a matter of

willing a maxim for exactly this occasion without taking it to have any other implications of any kind for any other occasion. You will a maxim thinking that you can use it just this once and then [...] discard it; you don’t even need a reason to change your mind (ibid., 75).

In order to establish yourself as an agent or a person, you must be able to distinguish *yourself* from the fleeting and unrelated impulses or motivational forces operating within you. This is impossible if you are merely a passive spectator to the battle between various desires and inclinations within you. Merely *observing* the workings of the different motivational states within you amounts to *particularistic willing*, namely regarding your final dominant desires as reasons applying this or that case and applying only for oneself. This means that *any* change in your motivational state would count as a reason against what you are doing at the moment. This is not making a decision or acting, according to Korsgaard.

Particularistic willing amounts to not willing at all because it lacks a subject. Willing an end conceptually involves a subject. This entails that particularistic willing is impossible. Thus, endorsing a desire and regarding it as having a (provisional) *universal*

normative force, i.e., being applicable to similar situations both for oneself and for others unless there is good reason why not, is constitutive of making a choice or acting or willing an end. For example, let's say I see two people fighting on the street and decide to intervene to make them stop. I will experience that decision as *my* decision only if I think that I would do the same thing regardless of my strongest desire to do otherwise (unless my strongest desire constitutes a good reason). That is, even if my strongest desire were to walk away because I did not want to ruin my jacket or to get hurt, *I*, as a psychologically unified agent, would still intervene because *I* would *endorse* my initial desire to intervene even in that case. Thus, on Korsgaard's view, *my choices* rather than my desires determine my actions, if my movements are to be called actions at all.

What are the standards under which a desire is a reason for action? In Korsgaard's view, standards of justification for endorsing a desire are laid out by one's *practical identity*, which is "a description under which you value yourself, a description under which you find your life to be worth living and your actions to be worth undertaking" (1996, 101). The *subjective* principles involved in one's particular practical identities constitute the standards according to which one decides whether to endorse a certain desire as a reason for action or not. This is the reason Korsgaard identifies action with *self-constitution*. For example, if I identify myself as a concert pianist, I will endorse my desire to practice the piano for four hours a day. If I identify myself as a social media influencer, I will reject my desire not to spend time on social media, and so on.

Korsgaard's third step is where she states her version of Kantian constitutivism: action constitutively involves trying to follow the Categorical Imperative. Korsgaard's argument so far, if successful, has established only that one must at least try to follow *subjective*

principles that apply universally throughout *one's own life*.⁵⁷ Korsgaard's account so far is compatible with immoral actions: one can act under the subjective principles of one's practical identity as a serial killer, for instance. So, for her theory to yield *moral* conclusions (or for constitutive standards for actions to have *categorical* normative significance), Korsgaard must show that action constitutively involves trying to follow universal principles that apply to *all* rational beings. Under the CI, one cannot consistently universalize the subjective principles endorsed by a serial killer. One cannot even universalize the principle (or maxim) of breaking a promise whenever it is convenient to do so if one is to follow the CI. So, Korsgaard may solve the problem if she can show that trying to conform to the CI is constitutive of action. But, again, to do that she must first show that action constitutively involves trying to conform to principles that apply to *all* rational beings rather than only to oneself.

Korsgaard is aware that her view so far is compatible with a form of egoism according to which one acts on reasons that apply universally throughout one's own life rather than on reasons that apply universally throughout everyone's lives.⁵⁸ To remove this complication, she makes a distinction between *private* (agent-relative) reasons and *public* (agent-neutral) reasons. A reason is private when it is normative just for *one* person, whereas a reason is public when it is normative for *all* people: "on the public conception of reasons, [...] if I have a reason to do action-A in circumstances-C, I must be able to *will* that you should do action-A in circumstances-C [...] because I must take your reasons for my own" (ibid., 191–2).

⁵⁷ If we take the meaning of 'universal' as being applicable to all cases, a universal principle can well be subjective.

⁵⁸ Korsgaard 1996, 134.

Korsgaard asserts that agents *must* regard reasons as public for interpersonal communication to be possible at all: “if personal interaction is to be possible, we must reason together [...] I must treat your reasons [...] as considerations that have normative force for *me* as well as you” (ibid., 192). Interpersonal communication involves “shared deliberation,” namely, thinking together and trying to arrive at a common decision about what to do based on a “shared good.” Interpersonal interactions also involve doing what one is told to do or refusing to do what one is told to do by proposing a reason for denial. For example, if my friend Joshua and I are trying to decide on a joint gift for our mutual friend’s birthday, I must regard the fact that Joshua does not like a certain brand of headphones as a reason for me to propose a different brand, or I must propose a reason to change Joshua’s mind about that particular brand. Or, if Joshua also thinks that headphones are not a good idea for a birthday gift, I must take his reasons as my own and propose something other than headphones. If I simply disregard his reasons and say “Well, headphones are good gifts. Let’s get the orange ones!” then we are not engaging in shared deliberation. If people disregard others’ reasons in this way, “no personal interaction is going to be possible” (ibid., 193).

Another possibility is that I might acknowledge the legitimacy of my friends’ having reasons for himself in the same way that I have reasons for myself, but I might at the same time see his reasons “only as obstacles to be defeated, or tools to be used” (ibid., 194) and try to manipulate him. According to Korsgaard, this would be “a kind of war,” not a shared deliberation. On Korsgaard’s view, if interpersonal communication were merely a combat between two or more private reasoners, then no reason would be normative for anyone. In other words, the fact that reasons are shareable with others

enables them to acquire normative status. This is Korsgaard's appeal to Wittgenstein's private language argument. Korsgaard does not infer her argument from Wittgenstein's argument but rather she draws an analogy: just as meanings of words cannot be normative unless they are shareable, so reasons cannot be normative unless they are shareable (ibid., 196n12). Shareability is the necessary condition for normativity.

If Korsgaard is correct about the publicity of reasons, then the universal principles we are trying to follow in order to act must be ones that we adopt as principles not only for ourselves but also for all rational agents. Since the CI is "the law of acting only on maxims that you can will to be universal laws [as opposed to subjective principles] (ibid., 80)," Korsgaard concludes that we must at least *try to conform* to the CI in order to act at all (ibid., 45, 81).

3.4.3 Problems with Korsgaard

We have seen in 3.3.3 that the appeal to inescapability may not be sufficient to defuse the dependency challenge to constitutivism. First, even if the question "Why be agents?" can only be asked from within agency, it might still be dialectically possible for an agent to question the normative force of being an agent and to challenge the binding force of the constitutive standards of agency. Or so Enoch (2006) believes. So, the constitutivist must show that there is *no* standpoint external to agency from which the skeptic can raise questions about the normative authority of agency's standards to defuse the dependency challenge. Otherwise, the constitutivist strategy cannot even get off the ground to secure objectivity and categorical normativity: one can always resort to stance-independent properties to explain the normative force of agency's standards. Second, even if agency is

inescapable and there is a categorical reason to value our own existence, it is still possible to ground the reason to mind one's self-loss in an absolute inner value that a realist would be happy to accept. So, the conditionality or dependency challenge should be met within the constitutivist framework. Otherwise, it is possible to ground the authority of agency or valuing in non-natural entities that realism postulates. This would be moral realism, not an alternative to it.

Korsgaard's constitutivism, I believe, can meet the first part of the dependency challenge. Korsgaard famously claims that "the necessity of choosing and acting [...] is our *plight*: the simple inexorable fact of the human condition" (Korsgaard 2009, 2).⁵⁹ That is, we cannot do something other than performing some action or another. True, we can suspend or terminate our own agency by sleeping or committing suicide. Agency is not *ontologically* inescapable.⁶⁰ However, even in those cases we are forced, by our very nature, to make decisions (to sleep or commit suicide) and implement our intentions to drop out of agency. We cannot exit agency *immediately*: exiting agency requires the exercise of one's agency. Thus, we unavoidably need reasons "to do one thing rather than another" (*ibid.*, 43): "*deciding* not to act is itself an action" (Katsafanas 2013, 52).

As Velleman (2004, 2009) and Ferrero (2010) claim, 'dialectical inescapability' is the best hope for the constitutivist to defuse Enoch's skeptical challenge because if agency is inescapable in Korsgaard's sense, then there is *no standpoint* outside of agency from

⁵⁹ Katsafanas, a Nietzschean constitutivist, agrees with Korsgaard: "action is inescapable: any attempt to avoid acting will itself be an action" (Katsafanas 2013, 53). For example, one must implement her intention to successfully commit suicide and drop out of agency. Without a doubt, this would be an exercise of agency.

⁶⁰ Cf. Ferrero 2018, 153.

which the skeptic could raise the question whether there is reason to be an agent.⁶¹ Any skeptical question about any activity (including questions about agency) can only be raised by subjects who are already exercising their capacity for rational agency. If the subject is already inside the practice of giving and asking for reasons, the question whether she has reason to opt into agency becomes unintelligible. This is why agency is “dialectically inescapable” and “closed under its distinctive operation” (Ferrero 2019, 12; 2010, 313).

There is only one standpoint whence the question whether there is reason to be an agent can be asked: the agent’s standpoint.⁶² An external standpoint is simply inconceivable. It is quite easy to raise the conditionality challenge with regard to ordinary activities such as games because one can easily occupy a standpoint outside of the activity in question. For example, I can ask whether there is reason to play baseball in the first place. When I ask this question, I am outside of the activity in question. Even a baseball player can occupy this external standpoint while playing baseball and question the normative force of playing baseball in the first place. The conditionality or dependency challenge is motivated by analogy with ordinary activities such as baseball. However, agency is crucially different from ordinary activities: it is dialectically inescapable, which is “a matter of the reflective closure of the exercise of practical reason” (Ferrero 2018, 127). Thus, the analogy breaks down: it is not even clear what the skeptic is asking about the binding force of the constitutive standards of agency (or about

⁶¹ Cf. Velleman 2004, 292–3; Ferrero 2010, 310–3; 2018, 127–8.

⁶² The kind of agency at stake is *intentional* or *rational* agency, which is exercised when one is not merely observing different motives within herself. (Recall Korsgaard’s argument against particularistic willing.)

our engagement in it) because we cannot even conceive ourselves outside of agency. This is precisely why Korsgaard says “constitutive standards meet skeptical challenges to their authority with ease” (2009, 29).

This seemingly trivial fact about agency is, in fact, an important achievement for constitutivism. The reflective closure of practical reason indicates that constitutivism should not give up its ambition to account for objectivity and categorical normativity simply due to the assumed possibility of an external standpoint from which one can question the normative authority of agency. If one cannot challenge the binding force of the constitutive standards of agency from an external standpoint, then the source of authority must lie within agency. This is good news for the constitutivist, who wants to ground authority in the constitutive standards of agency.

But does this really rule out realism? It seems that we can still ask intelligible questions about the normative force of agency from an *internal* standpoint. It may not be intelligible to ask whether one should opt into agency, but one can always ask whether she should remain to be an agent or whether she should exit agency. People can think about suicide. I can intelligibly ask whether I have a compelling reason to (continue to) be an agent. Nothing prevents me from asking this question *within* agency. So, internal normative challenge remains, even if the inescapability claim works against the skeptic. Couldn't it be that the source of authority for being an agent or for minding one's self-loss is an absolute inner value property? Dialectical inescapability of rational agency can at most establish that the source of moral objectivity and categorical normativity lies within practical reason. It does not by itself settle the question of how objectivity and the binding force of morality takes place, which is precisely what the constitutivist must

explain if it is to be a plausible alternative to moral realism. It is still possible that morality is grounded in a value property that is attached to our rational nature. Thus, Korsgaard's inescapability claim cannot rule out realism.

Moreover, the skeptic could claim that the power of the dependency/conditionality challenge lies not in the possibility of taking up an external standpoint but in the apparent gap between the Constitutive Claim (description of the constitutive features of agency) and Normative Claim (normative significance of agency). In other words, why merely descriptive features of agency have normative significance? Korsgaard takes morality to be constitutive of the agency-enabling principles that provide psychological unity. Why are these principles normatively significant?

As we have seen, Korsgaard's answer is that action constitutively involves a commitment to Kant's CI. If action constitutively involves trying to follow universal principles that apply to all rational beings, then complying with the constitutive standards of action must carry with it normative significance. Under the CI, one cannot consistently universalize the subjective principles endorsed by a serial killer, a devious manipulator, or a free-rider. And since the constitutive standards of action are *moral* standards, they have normative significance. Moreover, on Korsgaard's account, agents necessarily regard reasons as public or agent-neutral. Otherwise, interpersonal communication would not be possible. Korsgaard further claims that interpersonal communication is not a combat between two private reasoners who are trying to manipulate each other. Rather, interpersonal communication involves shared deliberation. The fact that reasons are shareable or communicable with others gives them a normative status: shareability is the necessary condition for normativity.

I will mention two further problems with Korsgaard account, the combined effect of which further explains why I don't prefer her version of constitutivism as an alternative to moral realism. First, even if Korsgaard could show that we must try to follow the CI to act at all, (a) she reduces moral normativity to the normativity of consistency, as FitzPatrick (2005) correctly points out, and (b) she –once again– fails to rule out realism. Second, Korsgaard cannot really establish that we must try to follow the CI to act at all. Any general subjective principle would enable us to exercise our agency.

Korsgaard's account is not that it is *literally* practically necessary to follow the CI in each particular case to act at all. Rather, on Korsgaard's view, we must at least *try* to follow the CI to act at all. That is, if Korsgaard's account of action is correct, we undertake a practical commitment to conform to the CI in acting at all, and if we do not conform to the CI in our particular actions, then these actions will contradict our practical commitment, which will amount to practical irrationality.⁶³ If it is not literally practically necessary to follow the CI to be able to act at all, then what explains the normativity of the CI? It perhaps would not make sense to question the authority of the CI if actually following it were literally practically necessary for action. But one still functions as an agent when he fails to conform to the CI. Korsgaard fails to explain the normativity of actually following the CI.

Korsgaard cannot resort to *literal* practical necessity when it comes to particular cases, since otherwise no immoral action would be possible. It is perfectly possible to be immoral and exercise agency at the same time. So, it must be that failing to conform to CI amounts to practical irrationality due to the inconsistency between our actions and

⁶³ Cf. FitzPatrick 2005, 672.

practical commitments. Moral normativity would then be reduced to the normativity of consistency. But why is consistency relevant to morality? Why is avoiding a violation of a practical commitment normatively more significant than following our desires? Is consistency between our practical commitments and our actions *intrinsically valuable*? But that would be realism. Korsgaard's constitutivism, once again, cannot rule out realism and fails to be an alternative to it.

On top of that, Korsgaard cannot establish that we must *try* to follow the CI to act at all. Korsgaard's argument against particularistic willing claims that in order to function as agents we must endorse desires as reasons for actions and we must regard our reasons for actions as having a *universal normative force*, i.e., our reasons must be applicable in similar ways to similar situations unless there is a good reason why not. So, in order to exercise agency, we must identify ourselves with general principles that reliably generate reasons for our actions. However, this does not entail that we must identify ourselves with the CI. Any general subjective principle would work. In fact, even the subjective principles involved in one's practical identity as a serial killer or a free-rider would be sufficient for one's movements to be called as actions. For example, a serial killer may think that everyone who suffers from loneliness has a reason to kill other people. (He might even make fun of others' inability to acknowledge –and act on– that reason.) When he acts on this subjective principle, he functions as an agent (although an immoral one).

Finally, shared deliberation still seems to be possible if our *private* reasons prompt us to cooperate with others and regard their reasons as normative. I may realize that I must cooperate with other to get them to do what I want or need, and I may act on this subjective principle. Moreover, Korsgaard thinks that all communication would be “a

kind of war, or combat” if reasons were private (2009, 194). But isn’t it possible for me to define myself as a good human being who always helps others and cooperates with them? This practical identity may involve the subjective principle that I should always deliberate *with* others rather than using their reasons merely as tools to get what I want or need. Isn’t it also possible for me to adopt others’ reasons as my own because of my belief that human nature has a non-natural value attached to it? I may think that all humans have this sort of dignity, and this belief could generate a private reason that could prompt me to take others’ reasons as my own. This process does not need to involve a conscious decision. I may simply find myself respecting others and ascribing normative force to their reasons. But as long as my taking others’ reasons as my own is due to my private reasons, it constitutes a counterexample to Korsgaard’s account. So, acting only on private reasons does not seem to entail that shared deliberation is impossible.

Even if we grant that the shareability (or publicity) of reasons makes them normative for all rational agents, we could still follow our private reasons that tell us to use others’ reasons as mere tools to advance our self-interest. On Korsgaard’s public conception of reasons, a consideration becomes a moral reason only if it can be shareable with others and consistently be universalized. But even if reasons were public in this way, this would only affect the nature of *moral* action, not necessarily of action *as such*. One would still be *acting* in using others’ reasons as “tools” to get what they want or need. Granted, on the public conception of reasons, one *should* adjust one’s private reasons such that they should take others’ reasons into account. But acting only on universalizable or shareable principles would not be necessary, although it would be the morally right option. One could act on their *immoral* subjective principles and still function as agents.

3.4.4 Transcendental Constitutivism

The above problems show that constitutivism must leave room for immoral actions. It is possible for one to act immorally and function as an agent at the same time. Thus, we must make a distinction between action and moral action. While subjective principles involved in one's practical identities may constitute action, the CI is supposed to constitute *moral* action. Here I present Sensen's reading of Kant, which supports this idea by associating moral value with how *pure* reason functions. This view is called 'transcendental constitutivism.'

I first explain why Sensen interprets Kant in this way. I then explain the metaphysical basis of such a view (3.4.5). Transcendental constitutivism adopts a Parfitian moral ontology. That is, on this view, moral facts exist in a *non-ontological* way just like mathematical and logical facts. I conclude the section with an account of how formal moral intuitions that I described in the second chapter (2.2.5) could be explained by how our reason necessarily functions. And theoretical intuitions could be partly explained by the application of *formal* intuitions to the *matter* provided by the details of one's emotional and social context (3.4.6). My claim is that transcendental constitutivism is a neglected alternative to moral realism.

I will begin with how Sensen's interpretation differs from that of Kantian realists. Kantian realism has been developed as a response to the traditional formalist reading of Kant's moral theory.⁶⁴ According to the formalist reading, there is *one and only*

⁶⁴ Cf. Sensen 2009, 262–3.

categorical imperative, which is the first normative reality: “*act only in accordance with that maxim through which you can at the same time will that it become a universal law*” (GMS 4:421). This imperative is a *direct* requirement of our reason and thus it is not grounded in any prior value. Thus, “*the concept of good and evil must not be determined before the moral law [...] but only [...] after it and by means of it*” (KpV 5:63), and “nothing can have a worth other than that which the law determines for it” (GMS 4:436). Kantian realists reject this reading and argue instead that Kant talks about an objective value property that constitutes a foundation for the CI.⁶⁵

The realist response is inspired by Hegel’s (1991, §135) famous ‘empty formalism’ charge against Kant, according to which Kant’s moral theory lacks content, authority, and motivation. The realist idea is that if one could show that Kant argues for an independent value that grounds the CI rather than defending the idea that the CI is given without reference to any value, then it would be easier to account for how concrete duties are generated, how the moral obligations are binding on us, and why we should be motivated to obey moral rules. Kantian realism, therefore, generally takes the form of a ‘value realism.’ That is, Kantian realists, as opposed to non-Kantian moral realism, do not argue for a stance-independent value that would have continued to exist if no human beings existed. Rather, Kantian realism claims that a value property is attached to our rational nature like an ontological diamond, i.e., if all human (or rational) beings died out overnight then no value is going to exist. The world would have been ‘morality-free’ if it did not contain any rational inhabitants.

⁶⁵ See Kain (2004, 2017), Langton (2007), Wood (2008), Hills (2008), and Schönecker (2017) for such a view.

It is important to note that different aspects of Kant's moral philosophy could be given different metaethical interpretations. For example, one could embrace one position about *value*, and another about *the moral law*. Or one could embrace one position about the *content* of the moral law, and another about the *authority* or *bindingness* of it.⁶⁶ Sensen's strategy proceeds in *three* main steps: (1) answering the question whether Kant holds that a metaphysical value property grounds the CI, i.e., whether Kant is a value realist; (2) answering the question, "What is the source of the *content* of the CI?"; and (3) answering the question, "What is the source of the *authority* of the CI?" After answering each question, he reaches the conclusion that Kant is a transcendental constitutivist about value, as well as the content and the authority of the moral law.

Kantian constructivists and Kantian value realists agree on Kant's point that morality arises from our own reason and not from anything external to it.⁶⁷ In other words, for Kantians, morality does not exist independently of the constitutive features of practical reason and what follows from them. This is the reason they agree that moral principles lie in the nature of practical reason.⁶⁸ Kantian value realists' point of departure is their assertion that a nonnatural value property that is constitutive of practical reason *precedes* the CI and *grounds* it. This idea is impelled principally by what Kant says about the unconditional value of a good will and the absolute value of humanity in the *Groundwork*. As regards the good will, the only thing that is good unconditionally, Kant

⁶⁶ Cf. Rauscher 2002, 485; Stern 2012, 68; Sensen 2013, 63; Formosa 2013, 173.

⁶⁷ Cf. KrV B1f.; KpV 5:31f.

⁶⁸ "The moral law [...] must be grounded [...] in the nature of practical reason or rational will" (Kain 2004, 290); "The content of the moral law lies in the nature (or essence) of the rational will or practical reason" (Wood 2008, 114).

asserts that even if it did not lead to any consequence or even if it is unable to carry out its purpose, “like a *jewel*, it would still shine by itself, as something that has its *full worth in itself*” (GMS 4:394f.; emphasis added). As regards the value of humanity or rational nature, Kant says that if “there were something *the existence of which has an absolute worth* [...] then in it, and in it alone, would lie the ground of a possible categorical imperative” (GMS 4:428; emphasis added). The thought that all human (or rational) beings have an absolute value is attractive because it makes it easier to accept Hegel’s objection yet reject that it undermines Kant’s position. On the value realist model, moral content or our concrete moral duties are ultimately provided by the value underlying the CI instead of the CI alone, which would be ungrounded and unable to provide moral content if rational nature had no ontological value property attached to it.

3.4.4.1 First Step: Kant against a Metaphysical Value Property that Grounds the CI

The first step for showing that Kant is a transcendental constitutivist is to show that Kant uses the term ‘value’ and the phrase ‘absolute value’ not to point to a metaphysical property a thing, person, or action possesses but to “express what reason dictates or prescribes under all circumstances” (Sensen 2013, 67). That is, Sensen takes value to be a secondary concept in Kant’s moral theory. He rests his interpretation on multiple textual sources of evidence.⁶⁹ I will mention only a few of them. First, Sensen talks about Kant’s arguments against the existence of a nonnatural or natural value property that grounds the CI. In the *Critique of Pure Reason*, for example, Kant gives an epistemic argument by claiming that one is not allowed to add non-natural entities to one’s moral ontology

⁶⁹ Cf. Sensen 2011, ch.1.

because there is no way in which we can discover them, i.e., we cannot discover such entities either by our senses or by intuition:

[W]e cannot originally cook up [...] a single object with any new and not empirically given property and ground a permissible hypothesis on it. [...] Thus, we are not allowed to think up any sort of new original forces, e.g., an understanding that is capable of intuiting its object without sense (KrV B798; cf. Sensen 2013, 70).

According to Sensen, Kant reinforces this point in the *Critique of Practical Reason* by asserting that the good cannot constitute the foundation for morality. Kant gives a *reductio* argument first by assuming that the good grounds morality and then by concluding that in such case “there would be nothing at all immediately good, and the good would have to be sought, instead, only in the means to something else, namely some agreeableness” (KpV 5:59).⁷⁰ That is, if the good were to precede and ground the CI, then the good would be external to the will and be “only good *for something*.” (ibid.) Since Kant thinks that a non-natural value property cannot be discovered by our senses or intuition, the only way we could determine what is good would be through a feeling of pleasure: “[i]f the concept of the good is not to be derived from an antecedent practical law but, instead, is to serve as its basis, it can be only the concept of something whose existence promises pleasure” (KpV 5:58). If we do not begin with the moral law (or the CI), the good “would be directed only to that which the feeling of *gratification* is immediately connected” (ibid.). And since pleasure is relative and contingent,⁷¹ it cannot

⁷⁰ “Suppose that we wanted to begin with the concept of the good in order to derive from it laws of the will: [...] the possibility of even thinking of a pure practical law [would] already [have been] removed” (KpV 5:63).

⁷¹ Cf. GMS 4:427, 4:441; KpV 5:58, 62.

be the source of the moral law.⁷² Sensen's conclusion is that Kant's epistemology refuses to postulate nonnatural value properties as the ground of the CI.

Sensen then argues that Kant does not conceive of value as a complex natural property such as pleasure or happiness either. First, as mentioned above, Kant argues frequently that morality cannot be based on pleasure. What causes pleasure differs greatly from one person to another; thus, pleasure is relative and contingent. Morality, according to Kant, is *a priori*, that is, it must be necessary and universal (KrV B4; GMS 4:389). Therefore, pleasure cannot be ground morality. Happiness cannot ground morality either, for Kant thinks that happiness is an indeterminate concept. In other words, one cannot consistently say what will make her happy, since happiness is an empirical concept that involves "a maximum of well-being in [her] present condition and in every future condition" (GMS 4:418). This means that if one wants to know which actions will make her happy, she must take into consideration infinitely many facts about her present and future states. It is, therefore, impossible to determine which actions are supposed to make one happy. Moreover, even if happiness were to be a determinate concept, it still could not ground morality because moral actions would not then be "objectively necessary of [themselves]" but would be relative to some further end, namely happiness (GMS 4:414). What makes people happy differs substantially among different people. Thus, happiness can only be "*subjectively necessary*" and can only yield hypothetical imperatives, which ground actions that are good "merely as a means *to something else*" (GMS 4:414; KpV 5:25). Kant excludes such imperatives from the realm of morality because they are "always conditional and could not be fit for a moral command" (GMS 4:433).

⁷² Cf. Sensen 2009, 270.

Another point Sensen makes is that Kant not only argues against a value property as the foundation for morality on epistemic grounds but also he never specifies value as a distinct property of objects, persons, or actions throughout his works (Sensen 2009, 267–268; 2011, 16–17; 2013, 68–69). First, according to Sensen, Kant does not conceive of value as a distinct property when he talks about possible moral theories. He mentions pleasure, happiness, perfection, divine command, moral feeling, custom, education, and civil constitution as candidates for value that grounds morality, but he never mentions a distinct metaphysical property (VE 29:620f.–629; KpV 5:39–41, 64). What is more, Kant claims that his list exhausts “all possible cases” for the “basis of morality.” (KpV 5:39; VE 29:620f.) This suggests that value as a distinct metaphysical property did not even strike Kant as a possible metaethical option.⁷³ Second, Sensen asserts that Kant does not base his justification of morality on a value property: both in the third section of the *Groundwork* and the second chapter of the *Critique of Practical Reason* Kant justifies morality without giving reference to any prior value (Sensen 2013, 69). Furthermore, in Sensen’s view, value does not play any role in Kant’s *Lectures on Ethics*, nor it is

⁷³ This should not be surprising because value as a distinct metaphysical property that objects possess inherently is a 20th century conception, popularized by G.E. Moore (1903). Moore’s conception of value can be contrasted with that of Plato. According to Plato, value cannot be found in the physical world, where objects and matter are merely imitations of nonphysical essences of things. Hence, objects and matter in the physical world are not as real as their nonphysical counterparts that Plato names ‘Forms.’ It follows that if morality is grounded in a nonnatural value it must lie outside of space and time, i.e., it must reside in what he calls ‘the world of Forms.’ (Plato 1997, 85–92, 93–95, 525, 1132–1137) Moore, on the other hand, is more like Aristotle in spirit, for on his account value is attached to our nature. However, according to Moore, the concept ‘good’ and any natural concept cannot have the same meaning. The property of goodness, therefore, cannot be identical to any natural property. He concludes that value or goodness is a stance-independent, nonnatural property that cannot be known empirically. Rather, we can have epistemic access to such a property only via our faculty of intuition (Moore 1903, ch.1).

included in his list of central moral concepts in the introduction to the *Metaphysics of Morals* (ibid.; cf. MS 6:221–228).

What about the good will and value of humanity passages in the *Groundwork*? Doesn't Kant talk about value as a distinct property that grounds morality there? After all, Kant says that even if a good will did not cause any action, "like a *jewel*, it would still shine by itself, as something that has its *full worth in itself*" (GMS 4:394f.; emphasis added), and that if "there were something *the existence of which has an absolute worth* [...] then in it, and in it alone, would lie the ground of a possible categorical imperative" (GMS 4:428; emphasis added). Sensen thinks Kant places emphasis on 'shine' rather than 'jewel' in the good will passage (Sensen 2011, 21; 2013, 72) That is, Kant uses the analogy to express the equality of relations. For instance, to say that "3 is to 9 as 4 is to 16" is not to say something about the inherent properties of 3 and 4. Rather, it is simply to say that 3 and 4 stand in the same relation to 9 and 16, respectively: 'be the square root of.' This indicates that Kant does not attribute any value property to a good will, but he simply asserts that a good will still *shine* like a jewel even when it does not bring about anything. That is, a good will demands *respect* from an observer.⁷⁴

Kant similarly does not evoke a metaphysical value property when he talks about the value of humanity, Sensen argues (2009, 270–271; 2011, 26–27). Sensen draws attention to the following passage: "if all worth were conditional and therefore contingent, then no

⁷⁴ "[B]efore a humble common man in whom I perceive uprightness of character in a higher degree than I am aware of in myself *my spirit bows*, whether I want it or whether I do not" (KpV 5:76f.–77). The kind of respect Kant talks about here refers to a feeling of esteem one has when she comes across a morally good will. According to Sensen, this kind of respect differs from the one that is shown to other people in general. The latter is to be understood as "the *maxim* of limiting our self-esteem [...] or] of not exalting oneself above others" (MS 6:449f.). See Sensen (2017, 208) for the two distinct kinds of respect Kant entertains in his moral theory.

supreme practical principle for reason could be found anywhere” (GMS 4:428).

According to him, this passage does not entail that there must be a value underlying the CI; rather, it is consistent with value being dependent on the CI. Consider the following *modus tollens*: “If there is a CI, there is something of absolute value. There is nothing of absolute value because all value is contingent. Thus, there is no CI” (Sensen 2011, 27).

The statement is, therefore, about whether the CI “*could be found* anywhere,” not about whether a value property grounds the CI. Moreover, the phrase “could be found” refers to an *epistemic* relation or the “*ratio cognoscendi*,” rather than an ontological one, the “*ratio essendi*” (KpV 5:4n.). That is, if we were never aware of the requirement that we must value humanity regardless of our desires, we would never be aware of the CI. ‘Absolute value’ is the *ratio cognoscendi* of the CI, while the CI is the *ratio essendi* of ‘absolute value.’

Sensen also takes Kant to hold that the ground of the CI is freedom in a *descriptive* sense (2009, 271). According to Kant, “[t]he ground of this principle [the CI] is: *rational nature exists as an end in itself*” (GMS 4:428f.). Kant identifies an ‘end in itself’ with freedom in a descriptive sense: “[i]t is freedom, and freedom alone, which makes that we are ends in ourselves. [...] If [one] is not free, [one] has to be in someone else’s hand, consequently always the end of someone else, and therefore a mere means” (NF 27:1322). In other words, one is free if she is not being pushed around by natural forces (e.g., impulses, desires, and so on), i.e., if she is not a means to the ends of nature.

Freedom in this sense does not have any normative implications but it merely *describes* how a fully or purely rational being would act.⁷⁵ Thus, to say that the CI is grounded in

⁷⁵ A fully, purely, or perfectly rational being always behaves in accordance with the dictates of reason, and thus its “volition is of itself necessarily in accord with the law: the ‘ought’ is out of

an ‘end in itself’ is to say that the CI is grounded in our ability to act according to what reason demands, i.e., to act independently of our desires and inclinations.⁷⁶ This is why Kant declares freedom – in the sense described above – to be the “*ratio essendi* of the moral law” (KpV 5:4n.). Kant does not invoke a distinct value property to ground the CI.

Sensen’s conclusion is that Kant uses the phrase ‘absolute value’ not to describe a value property that grounds morality but merely to express what is practically necessary: “the will is a capacity to choose *only that* which reason independently of inclination cognizes as *practically necessary*, that is, as *good*” (GMS 4:414; emphasis added). In Kant’s moral philosophy, value is not a property an object possesses inherently but it refers to an action or person:⁷⁷ (i) an action has absolute value if reason prescribes it under all circumstances (KpV 5:60), (ii) a person has absolute value if she *always* chooses what is practically necessary.⁷⁸ ‘Absolute value’ is merely a shorthand for *what*

place” for such a being (GMS 4:414). The moral law becomes normative for rational beings who are also affected by desires and inclinations: “the law has the form of an imperative [for human beings, since they are] affected by needs and sensible motives” (KpV 5:32). For Kant, then, freedom is *not a normative* concept but a *descriptive* one.

⁷⁶ Cf. GMS 4:446–447; KpV 5:29; KrV B476, B562f.

⁷⁷ “Thus good or evil is, strictly speaking, referred to actions [...] and if anything is to be good or evil absolutely [...] it would be only the way of acting, the maxim of the will, and consequently the acting person himself [...] but not a thing” (KpV 5:60).

⁷⁸ In Sensen’s interpretation, Kant does not think that every human being has a good will solely in virtue of their rational nature. When Kant talks about a vicious person who nevertheless deserves respect, he says that “I cannot withdraw at least the respect that belongs to him in his quality as a human being, *even though by his deeds he makes himself unworthy of it*” (MS 6:463; emphasis added). That is, we must respect even a vicious person not because he has an intrinsic value property but because it is commanded by reason, i.e., *it is our duty*: “to deny them the respect owed to human beings in general, is in every case contrary to duty,” which is the “practical unconditional necessity of action” and “can be expressed only in categorical imperatives” (ibid.; GMS 4:425).

should be valued under all circumstances. The postulation of a distinct value property is not required to ground morality.

3.4.4.2 Second Step: The Source of the Content of the Moral Law

Sensen's second step for showing that Kant is a transcendental constitutivist is to show the *source* of the moral law. That involves first showing the source of the *content* of the moral law and then the source of the *authority* of it. As regards the source of the content of the moral law, Sensen claims that "the moral law is a direct command of reason" (2013, 73). This means that reason gives of itself the moral law rather than mirroring a separate moral reality: "[p]ure reason [...] gives (to the human being) a universal law which we call the *moral law*" (KpV 5:31f.). Reason provides the content of the moral law *directly and spontaneously*, meaning that it does not discover the law in experience or by intuition, but it gives the law "out of itself:" "pure reason, *practical of itself*, is here immediately lawgiving" (KrV B1f.; KpV 5:31). When Kant talks about reason as the source of the moral law, he does not refer to conscious deliberation. He rather refers to *pure* reason, "in which no experience or sensation at all is mixed in" (KrV B24).

Sensen makes three claims about the content of the moral law: (1) it is not under one's voluntary control, (2) it does not change, and (3) it is laid out prior to one's conscious awareness (2013, 74). First, one cannot change the content of the law at will as one could change the rules of etiquette. Kant makes this very clear with his distinction between legislator and author:

The law giver [or legislator] is not always simultaneously an originator [or author] of the law; [...] if the laws are practically necessary, and he merely declares that they conform to his will, then he is a lawgiver. So nobody, even the deity, is an

originator of moral laws, since they have not arisen from choice, but are practically necessary (VE 27:283).

In Kant's view, then, the author determines the content of moral obligations, whereas the legislator chooses maxims that are compatible with these obligations. And since even God cannot be the author of the moral law, we cannot change the content of the law by our actual acts of willing.

Second, the content of the moral law is outside of space and time: “[p]ure reason [...] is not subject to the form of time [...] and] no temporal sequence takes place in it even as to its causality” (KrV B579, B581). Since the concept of change presupposes time, reason “does not alter” (KrV B584), nor does the content of the moral law it produces.⁷⁹

Third, pure reason gives the law to us prior to our conscious awareness. For Kant, conscious deliberation is always in time and thus in the physical world (KrV B152–3), whereas the moral law and its source, namely pure reason, are outside of time and are “distinct from all determining grounds of events in nature” (KpV 5:28f.).⁸⁰ Thus, “the mere form of a law can be represented only by reason,” and not by conscious deliberation (KpV 5:28).⁸¹

Proving the above three claims to be true does not necessarily show that Kant is a constitutivist about the content of the moral law. One must further show that the moral

⁷⁹ This point constitutes an objection to transcendental constitutivism: how is it less mysterious than moral realism? I will address this objection in 3.5.1.

⁸⁰ In 3.5.1, I explain how pure reason can nevertheless exist in the physical world as an emergent, unalterable structure of thinking.

⁸¹ Even though the moral law is not the product of our conscious deliberation, “[w]e can become aware of [it ...] by attending to the necessity with which reason prescribes them to us and to the setting aside of all empirical conditions to which reason directs us” (KpV 5:30).

law is *not* an innate principle to distinguish Kant's position from moral realism.⁸²

Morality could be innate in the sense that our sense of moral duty and particular moral requirements have been embedded in the genetic code of human beings before we start making decisions. (Recall Joyce's evolutionary story.) That is, we might have been evolved in such a way to respond to an independent moral reality. Or the moral law might have been placed within us by a higher power for us to be able to be governed by an independent moral reality. In both cases, morality would be something to be discovered.

According to Sensen, Kant does not conceive of the moral law as an innate principle because, otherwise, it would not be necessary. Kant claims that the moral law is *a priori* (KpV 5:31). And when Kant talks about an *a priori* principle, he means that the principle is "not based on any intuition, either pure or empirical," and that it is necessary and universal (ibid.).⁸³ For example, when Kant describes the categories of the understanding, which are *a priori* elements of cognition that necessarily structure our sense experience, he rejects that the categories are "subjective predispositions for thinking, implanted in us along with our existence by our author" (KrV B167). Otherwise, "the categories would lack the **necessity** that is essential to their concept." Moreover, the moral law could not have been acquired by evolutionary processes either because "we must not let ourselves think of wanting to derive the [moral law ...] from the *special property of human nature* [... or] from the special natural constitution of humanity" in order for moral principles to "hold for all rational beings" (GMS 4:425).

⁸² Cf. Sensen 2013, 75–76; 2017, 201–203.

⁸³ "Necessity and strict universality are [...] secure indications of an *a priori* cognition, and also belong together inseparably" (KrV B4).

A priori principles, according to Kant, are not based on a “subjective necessity [that is] arbitrarily implanted in us,” but rather they are “originally acquired” (ÜE 8:222). Categories, for example, do not exist prior to our representations of objects in the world. Rather, once one is confronted with unprocessed sense-data, the understanding spontaneously provides the categories out of itself and thereby structures our sense experience (ÜE 8:221). Likewise, we become conscious of the moral law “as soon as we draw up maxims of the will for ourselves” (KpV 5:29f.). That is, once we start deliberating about morality, our reason “with complete spontaneity [...] makes its own order according to ideas [...] it even declares actions to be necessary” (KrV B576). So, the content of the moral law is not innate, but it is the result of a “spontaneous activity of reason:” “it is a principle that is constitutive of how pure reason operates” (Sensen 2013, 77; 2017, 218).

This is where the view Sensen attributes to Kant differs from Korsgaard’s constitutivism. While on Korsgaard’s account the CI is constitutive of empirical willing, on Sensen’s account it is constitutive of *pure* reason (Sensen 2017, 218–9). According to Korsgaard’s constitutivism, in order for us to be able to act or choose anything at all, we must try to follow the CI. On Sensen’s model, an agent can deliberately choose to perform immoral actions; however, in order to be *free* in the sense described above, she *must* follow the CI. Thus, Sensen concludes that Kant is a *transcendental* constitutivist about the content of the moral law.

3.4.4.3 Third Step: The Source of the Authority of the Moral Law

Sensen then moves on to the source of moral normativity: what grounds the *rational authority* of the moral law? (2013, 80) First, in Kant's view, the source of moral obligation is not something that is external to the will, such as desires or a stance-independent value property, but the will of every rational being must *bind itself* by being a law to itself. The will gives the moral law out of itself and *for itself* (GMS 4:440–5). Kant identifies the source of moral normativity with the 'law-giver:' "the term 'law-giver' [...] should designate only that man who is *necessitator*, in order to determine the will to observance of a law which the other knows" (VE 27:546). The law-giver or necessitator *cannot* be another will; otherwise, the moral law would only be *conditionally* binding: "the one imposing obligation [...] could always release the one put under obligation [...] from the obligation" (MS 6:417). For the moral law to be *unconditionally* binding, the law-giver must be "only *our own will*, insofar as we make it general, and regard it as a universal law" (VE 29:627; emphasis added). The source of the authority of the moral law is then the fact that the moral law is provided by one's own will or reason: "if the duty is a duty to myself, I think of myself as *binding* and so as actively constraining (I, the same subject, am imposing obligation)" (MS 6:417). Therefore, Sensen concludes that Kant is a transcendental constitutivist about the *authority* or *bindingness* of the moral law.

To sum up, Sensen's Constitutive Claim is that the moral law (or the CI), which provides reasons for action, is a constitutive feature of *pure* reason. His Normative Claim is that the moral law is unconditionally binding on human (or rational) beings because it arises from the spontaneous activity of *our own reason*. And his Content Claim is that the

content of first-order moral rules is extracted both from one's desires and inclinations, which determine one's subjective reasons for action (or maxims), and from the CI, which determines one's duty. To determine moral truth, one must reflect on one's desires and inclinations from the stance of pure reason. Since desires and inclinations are a part of the answer to the first-order question of what moral truths there are, Sensen believes that Kant is a *constructivist* about normative ethics (2013, 81).

Transcendental constitutivism differs from Kantian value realism because it denies the existence of a metaphysical value property as the foundation for morality. It differs from Korsgaard's constitutivism because it takes the CI to be constitutive of *pure* reason (*Wille*) rather than empirical willing (*Willkür*). And it differs from ideal stance constructivism because transcendental constitutivism identifies moral goodness with what reason deems necessary "*under all circumstances*" (Sensen 2013, 67; emphasis added) rather than with "the *desires* any person would have *under conditions of full rationality*" (Smith 1994, 199; emphasis added). While ideal observer theories conceive of a fully rational being as a fully informed and coherent agent with desires, transcendental constitutivism claims that a fully rational being is the one who does *not* have any desires or inclinations.

3.4.5 Parfitian Moral Ontology

Transcendental constitutivism describes moral value as a constitutive feature of how pure reason operates. But how should we conceive of moral value? What is the metaphysical basis of transcendental constitutivism? This view is an alternative to moral realism because it does not refer to a stance-independent value property. We have seen in the first

chapter that ontological characterization of moral value and perceptual characterization of moral knowledge creates metaphysical and epistemological problems for moral realism. I also claim in the first chapter that drawing an analogy between empirical facts and moral facts runs the risk of conflating the kind of objectivity possessed by empirical facts with the kind of objectivity possessed by moral facts. This is because empirical and moral facts are fundamentally different. While empirical facts are about ‘*what is the case,*’ moral facts are about ‘*what ought to be done.*’ Given this fundamental difference, it would not be very surprising if it turned out that moral facts had a different metaphysical status than empirical facts, and that they possess different kinds of objectivity. How should we explain that difference in metaphysical status?

This fundamental difference between empirical and moral facts has prompted some philosophers to adopt a *quietist* view about moral ontology. Quietists argue that moral and empirical facts are different, but they do not explain *how*. For example, Scanlon (2014, lecture 2) says we can vindicate moral reasons from within the moral domain, without the need to appeal to a domain-general ontology regarding what really exists. In other words, we can engage in moral reasoning and settle moral issues without worrying about how morality fits with metaphysical commitments of other disciplines or domains.

Scanlon grants that moral judgments have ontological implications, but his distinctive claim is that moral ontology is *domain-specific*, which means that the best reasoning within the moral domain determines what exists in that domain. Just as good mathematical reasoning gives us conclusions such as “The only even prime number is 2,” so good moral reasoning gives us conclusions such as “Racial discrimination is wrong.” Numbers and mathematical truths exist, but they are not located in the world that is

described by the sciences. Similarly, moral reasons and moral properties exist, but ontological commitments of moral judgments do not have anything to do with what the physical world involves. Recall Nagel's theoretical conception of morality that has its internal standards and methods of justification (2.2.4.2). In Scanlon's view, we can derive ontological facts from standards and methodologies of domain-specific discourses such as moral discourse, as long as these facts are internally coherent and do not directly contradict scientific facts. Otherwise, scientific ontology does not enjoy a privileged status over moral ontology or any domain-specific ontology.

The quietist view is motivated by the fact that certain domain-specific claims do not require a second-order ontological inquiry into what "really" exist. Consider mathematical or logical reasoning. When mathematicians or logicians engage in (first-order) mathematical or logical reasoning, they only need to apply appropriate domain-specific standards and methods. They do not also need to determine the ontological nature of mathematical or logical truths, or how they fit with the physical world. Similarly, when philosophers engage in moral reasoning, they do not need to also figure out whether moral facts "really" exist. This is an intuitive idea: no ontological inquiry is needed to establish that Jim Crow laws are morally wrong because they are based on racial discrimination.

This, I believe, is what Scanlon gets right: we do not need to find a way of locating moral reasons outside of the moral domain. However, Scanlon's claim that each domain-specific discourse, especially moral discourse, has ontological implications makes his position problematic. If ontological facts are determined by first-order domain-specific reasoning, then any claim from within a domain-specific discourse will get an ontological

free pass. For example, different descriptions of God from within different theological domains will all have ontological implications, as long as they are internally coherent and do not directly contradict scientific facts. It is not difficult, for example, to imagine an intervening God that is compatible with evolutionary theory, but it is far from certain whether such a God exists in the actualist sense.⁸⁴ FitzPatrick (2016) rightly calls this the “proliferation problem.”

The solution may be to take quietism further than Scanlon and claim that moral truths do *not* have any ontological implications. This is how Parfit (2011b, ch. 31; appendix J) describes moral ontology. According to Parfit’s non-metaphysical cognitivism, “even if nothing had ever existed, there would have been prime numbers greater than 100” (2011b, 21). Likewise, even if nothing had ever existed, there would have been moral principles. Just as numbers, mathematical or logical truths do not exist in the actualist sense, so the moral truths (or to be more precise, moral aspects of actions) cannot be found in the physical world. Rather, mathematical, logical, or moral truths *exist* in a *non-ontological* sense.

In Parfit’s view, “exists” or “there is” has different senses. Anything that is logically possible or conceivable exists in the *wide*, general sense. Merely possible things exist *only* in this wide sense, but they have lesser ontological status than things that *actually* exist. Physical objects and physical aspects of actions (or anything that the empirical sciences describe) exist both in this wide sense *and also* in the narrow actualist sense. Abstract things such as mathematical facts, logical facts, and normative/moral aspects of actions/persons *exist* both in the wide sense and in the *non-ontological* sense. As opposed

⁸⁴ Actualism is the view that everything there is exists or is actual.

to merely possible things, however, things that exist in this non-ontological sense do not have lesser ontological status than things that actually exist. Since mathematical, logical, and normative truths are *necessary* truths, “they do not have to be made true by there being some part of reality to which they correspond [...] It is reality that must correspond to these truths.” This is an intuitive idea. We engage in mathematical, logical, and moral reasoning, which yield certain conclusions, and we shape our environment, thoughts, or actions according to these conclusions. It is not that empirical facts determine mathematical, logical, or normative facts.

Is Parfit’s moral ontology the metaphysical basis of transcendental constitutivism? I believe it is. According to Parfit,

truths do not have to exist, or be real, in an ontological sense. Truths need only be true. [...] There are some claims that are irreducibly normative in the reason-involving sense, and are in a strong sense true. These truths have no ontological implications. For such claims to be true, it need not be true that reason-involving properties exist either as natural properties in the spatio-temporal world, or in some non-spatio-temporal part of reality (2011b, 21).

There are striking similarities between Parfit’s non-metaphysical cognitivism and Kant’s moral ontology. According to Kant, morality can be “firm even though there is nothing in heaven or on earth from which it depends or on which it is based” (GMS 4:425). Kant’s moral ontology is not based on natural or non-natural moral properties:

such a completely isolated metaphysics of morals, mixed with no anthropology, theology, physics, or hyperphysics and still less with occult qualities (which could be called hypophysical), is [...] an indispensable substratum of all [...] cognition of duties (GMS 4:410).

So, according to the transcendental constitutivist reading of Kant, moral truth is not determined by natural or non-natural properties but by the principle and commands provided by pure reason. Kant is not a moral nihilist, so he thinks that there are moral

truths. It follows that what he sees as the source of moral truths, namely pure reason, must also exist. But since Kant does not base morality on any natural or non-natural property, it seems that he regards moral truths as *existing* in a non-ontological way, similar to mathematical and logical truths. I thus believe transcendental constitutivist reading of Kant adopts a Parfitian moral ontology.

This does not mean that Parfit himself is a transcendental constitutivist. In fact, Parfit is a realist. He believes that mathematical, logical, or moral principles are *stance-independent*: “reason-involving normative properties, are, I believe, of this independent kind” (Parfit 2011b, 475). Interestingly, since he denies that moral properties exist in an ontological way, some philosophers are reluctant to call him a realist. FitzPatrick finds it “awkward to describe Parfit as a normative realist.” (2016, 539) Mark van Roojen agrees: “insofar as Parfit would seem to deny any ontological commitment to properties, I don’t know where to place him” (2015, 276).

Granted, a commitment to a non-ontological existence circumvents metaphysical objections: “when some view has no metaphysical implications, it cannot be open to metaphysical objections” (Parfit 2011b, 747). This, in fact, what motivates me to talk about Parfit’s moral ontology in the first place. Nevertheless, some kind of mystery will remain if we are going to say that these non-ontologically existing things are independent of us. As FitzPatrick rightly complains,

what the non-metaphysical view asks us to accept is that there are these irreducibly evaluative or normative properties and facts but that they are no part of reality, instead floating free of the world even as they are about worldly things [...]; they are not made to obtain by anything in the world, being modeled instead on abstract truths of logic or mathematics [...], and yet they have the deepest practical significance for how we should live (FitzPatrick 2016, 541).

What creates the mystery, I believe, is that Parfit overlooks the possibility that the source of morality might be *our own reason*, as transcendental constitutivism has it. Once again, the problem is stance-independence. If moral principles are provided by our own reason, then it is more likely that they will have the deepest practical significance for how we should live. Moreover, it is far from clear how FitzPatrick's bloated moral ontology itself can account for the practical significance and the authority of morality, as he conceives of moral properties as being independent of human reason (recall the discussion in 1.4.3). We can ask the same question to FitzPatrick: why would irreducibly normative properties that are *independent* of our rational capacities have the deepest practical significance for how we should live? This discussion shows further that transcendental constitutivism is a neglected alternative to moral realism. Parfit provides a metaphysical view that is more exhaustive than actualism, but he overlooks the transcendental constitutivist alternative when locating the source of morality.

Admittedly, transcendental constitutivism (or constitutivism in general) cannot *defeat* the skeptic who questions the practical significance or the categorical authority of morality, although the appeal to dialectical inescapability can defuse her by showing that the source of authority must lie within practical reason (recall the discussion in 3.4.3). The skeptic can always ask questions such as, why care about what my reason commands? This internal question, however, may never be answered satisfactorily by any metaethical theory, not only by transcendental constitutivism.

Transcendental constitutivism, as the name suggests, uses a transcendental argument to locate the source of morality: *If* morality is objective and categorically normative, and if we *can* act morally, then we must be *free* in the sense that we must be able to act

against our strongest desires and inclinations. This sort of freedom (pure reason or *Wille*) is the metaphysical source of morality (even though the generation of morality from that capacity is to be viewed in a non-ontological sense).⁸⁵

Transcendental arguments, however, have a limitation in that they cannot *refute* skepticism. First, one may claim that morality is *not* objective or categorically normative to begin with. Second, one may deny that pure reason is the metaphysical source of morality even if morality is objective and categorically normative. Third, one may accept transcendental constitutivism but still question the *authority* of the moral principle that is laid out by her own reason; that is, one may always choose her desires over moral requirements. And fourth, it is not logically impossible that there are beings with radically different cognitive faculties and conceptual schemes, who can ask *external* questions about the authority of morality.

The possibility of these four kinds of skepticism does not pose a distinctive problem for transcendental constitutivism because *no* metaethical position can refute these kinds of skepticism. No metaethical position can eliminate the possibility of indifference or the possibility of beings with a radically different mind or cognition. The possibility of skepticism is due to our own cognitive limitations, not due to the weakness of transcendental constitutivism. Transcendental constitutivism, unlike realism, ties morality to our own reason, which makes it more likely that morality is binding on us. And since it adopts a Parfitian moral ontology, it does not suffer from the problem of explaining supervenience.

⁸⁵ More on this in 3.5.1 and 3.5.3.

3.4.6 Moral Intuitions

Complementing Parfit's moral ontology with a transcendental constitutivist, *stance-dependent* origins story may have *epistemic* advantages as well. If the moral principle, the CI, is constitutive of our reason (just like the categories of the understanding), then it may become easier to explain how we can attain moral knowledge. Just as the categories of the understanding (the a priori elements of cognition) necessarily structure our sense experience, or, just as logical and mathematical principles necessarily structure how we think theoretically, so the CI necessarily structures how we think practically.

For example, if logical laws govern our theoretical thinking, then we must apply the rules of logic to be able to think at all. It is not as if we must first learn the rules of logic to be able to engage in theoretical reasoning: even a very young child can understand what a contradiction is. Similarly, if the CI governs our practical reasoning, then we must apply it to be able to reason practically at all. It is not as if we must first learn moral rules to be able to engage in moral reasoning: even a very young child can make a distinction between moral rules and mere conventions.⁸⁶

If transcendental constitutivism gets things right; that is, if we have no option but to think in terms of the CI while engaging in moral reasoning, then it may be easier for us to reach moral knowledge than the picture presented by moral realism. It may not be possible to know whether we are really free in the sense of having the ability to act against our strongest desires and inclinations due to our cognitive limitations. If we are not free in this sense, then morality may not be objective or categorically normative after all. However, if moral laws structure our thinking just as logical or mathematical do, then

⁸⁶ See Nichols 2004 for an experiment that confirms this.

we are already thinking in terms of the CI when we engage in moral reasoning, and we can apply this formal principle to particular cases to derive specific duties. That is, we can apply the CI to our emotional and social-environmental context to determine more specific moral content. Compared to the realist story, according to which we are trying to discover something that is independent of us, transcendental constitutivism arguably gives us a more appealing epistemic picture.

Transcendental constitutivism is also compatible with the account of moral intuitions I described in the second chapter. To reiterate, on that account,

(1) Actions arouse emotions. Due to our shared biological nature, we are inclined to react in particular ways to particular actions. These reactive attitudes are the products of our ‘*concrete intuitions*.’ For example, when we hear about an instance of incest, or someone beating his child, or someone cheating on their partner, our concrete intuitions make us feel that they are wrong, and we react accordingly.

(2) We use our capacity for abstraction (the capacity to see actions or objects as members of general categories) and find certain *kinds* of actions right or wrong (e.g., “Incest is wrong”). We call such reactions ‘*mid-level intuitions*.’ Mid-level intuitions are similar to what Haidt (2001, 828) calls “*post hoc* rationalizations:” They appear to originate from moral reasoning, but they are often merely expressions of our emotions (plus abstraction).

(3) We systematically think about our concrete and mid-level intuitions and question their appropriateness, credibility, coherence with each other, and so on. We reflect on various circumstances we might find ourselves in and whether different conditions affect the rightness/wrongness of particular actions. We also try to see whether our concrete and mid-level intuitions check with our formal intuitions. After reflecting systematically on our concrete and mid-level intuitions, we reach generalizations or abstract moral theories (*non-formal theoretical intuitions*). And when we can think of exceptions or counterexamples to our moral theories, we either reject them completely or revise them.

(4) *Formal intuitions* place formal constraints on our moral judgments and moral theories rather than making a moral evaluation. Formal intuitions are entailed by the nature of evaluative/normative/moral concepts, such as ‘better than,’ ‘wrongness,’ ‘permissibility,’ ‘reason,’ and so on. They are also governed by logical principles such as the principle of non-contradiction. While reasoning from concrete cases to theoretical intuitions is *inductive*, the reasoning from formal intuitions to moral facts is *deductive*. Formal intuitions function in a similar way to axioms in geometry: we derive specific moral content from them.

Formal moral intuitions govern practical reasoning, and they could be explained by *a priori* elements that lie within pure reason. Practical reasoning is governed by principles such as

- (a) If X is better than Y and Y is better than Z , then X is better than Z .
- (b) If it is wrong to do X , and it is wrong to do Y , then it is wrong to do both X and Y .
- (c) If two states of affairs, X and Y , are so related that Y can be produced by adding something valuable to X , without creating anything bad, lowering the value of anything in X , or removing anything of value from X , then Y is better than X .
- (d) If X is a reason for Y , it is not the case that that X is not a reason for Y .
- (e) If only facts of kind X are reasons for Y , and Z is not of kind X , then Z is not a reason for Y .
- (f) If you have a reason to do X , you also have a reason to take what you acknowledge to be the necessary means to X .

Formal intuitions are the products of the constitutive features of evaluative or moral concepts such as ‘being better than,’ ‘wrongness,’ or ‘being a reason for something.’ People with different levels of education, different (ethical) upbringings, different trainings, and different life experiences may hold different and conflicting normative or metanormative views. For example, some people may claim that moral and even logical principles are merely products of our social embodiment. However, even when these people are making these claims, formal intuitions may already be operating in the background, which is something they cannot change simply by making claims about social embodiment. *All* people, simply by virtue of their rational nature, may *necessarily* be thinking in terms of these formal intuitions, regardless of the conclusions they could reach due to different upbringings, trainings, experiences, and so on. This may be because formal intuitions, just as the CI, are grounded in *a priori* elements that lie within pure reason, as transcendental constitutivism would have it. On this view, formal moral

intuitions such as the ones listed above govern practical reasoning just as the CI (the universalizability principle) does. So, we could add the principle, “*act only in accordance with that maxim through which you can at the same time will that it become a universal law,*” to the above list (GMS 4:421). These formal intuitions, just as the CI, could inform our concrete intuitions and our judgments about particular cases, and they could all have their source in pure reason. Since these intuitions would arise from our own reason, the talk of intuitions in this sense would not support rational intuitionism of realism but rather it would support transcendental constitutivism.

Could formal principles that are constitutive of moral concepts or the moral principle (the CI) itself change? According to transcendental constitutivism, they cannot change because what follows from pure reason is not subject to time: reason “does not alter” (KrV B584). Nevertheless, it is logically possible that there are beings with radically different cognitions and conceptual schemes than humans. Such beings’ thoughts and actions may be governed by different concepts than ours. They may have a radically different picture of what it is to be a reason, what it is to value, what it is to act, and so on. Or they may simply lack those concepts. But we simply cannot understand what such a picture amounts to. For example, what does it mean to say that one *acts* intentionally without at the same time taking something to be a reason? What does it mean to say that “X is a reason for Y and X is not a reason for Y”? These are simply unintelligible for beings like us. Formal intuitions may, therefore, reflect our cognitive limitations. They draw the line between what is intelligible and what is not. I believe transcendental constitutivism would agree with this.

Does this view conflate practical and theoretical principles? The above presentation emphasizes certain logical and mathematical principles such as the law of non-contradiction, the principle of additivity, and the transitive law and ties them to practical reasoning. But didn't I make a distinction between the theoretical and the practical when I claimed that logical and mathematical principles govern theoretical reasoning, whereas the CI governs practical reasoning? Not necessarily. I made that claim to show the similarity of *function* between logical/mathematical principles and the moral principle (the CI): they all govern our thinking. They are all constitutive features of pure reason, and they work together in governing our reasoning, theoretical *and* practical. In fact, according to Kant, there is no real difference between theoretical and practical reason: "if pure reason of itself can be and really is practical, as the consciousness of the moral law proves it to be, it is still only one and the same reason which, whether from a theoretical or a practical perspective, judges according to a priori principles" (KpV 5:121). So, reason is one and the same in Kant's philosophy. We sometimes apply reason to the realm of "what is," and we sometimes apply it to the realm of "what ought to be." But in both of reason's applications (theoretical and practical), reason refers to the ability to act according to principles. According to transcendental constitutivism, our knowledge (both theoretical and practical) is shaped by the formal constraints that we necessarily have as rational agents. My claim is not that transcendental constitutivism is correct. Rather, my claim is simply that transcendental constitutivism is a neglected alternative to moral realism.

3.5 Objections to Transcendental Constitutivism

3.5.1 Appeal to a Noumenal Realm

Transcendental constitutivism, just as Korsgaard's constitutivism, tries to capture the objectivity and categorical normativity of morality by identifying moral value with a particular *function* of our reason. However, instead of reducing value to the facts about human psychology, transcendental constitutivism identifies moral value with the workings of *pure* reason, which is "cleansed of everything empirical" (GMS 4:388f.). This might raise the question whether this view is subject to Mackie's queerness argument or whether it appeals to a noumenal realm to which we have no access. To distinguish its non-naturalism from that of moral realism, transcendental constitutivism should first address this objection.

First, transcendental constitutivism does not necessarily commit itself to mysterious non-natural entities or to a noumenal realm because a process in nature such as reason can be recognized as possessing necessity, but the justification for its necessity may require the non-natural methodology of the transcendental argument, which does not necessarily have any ontological implications. That is, if we distinguish between the *metaphysical status* of a process in nature and *methodological justification* of it, we do not need to worry about the queerness argument (or evolutionary processes).⁸⁷ For example, the laws of aerodynamics determine the conditions of flight, to which organisms or objects must conform in order to be able to fly. Birds have been evolved in a way that they acquired the ability to fly, but the conditions of flight would still have

⁸⁷ Cf. Rauscher 2006.

been valid even if no flying organism evolved. Similarly, once organisms evolve in a way that they embody “the conditions of reason,” their mind starts functioning in a way that places formal restrictions on action. However, the *validity* of the conditions of reason is justified *a priori*, that is, independently of how organisms evolve in nature. According to this interpretation, then, there is *no* separate pure reason that is “cleansed of everything empirical,” existing independently of certain organisms that manifest the conditions of reason. That is, it is possible to use Kant’s transcendental language as a methodological tool to explain our experience of morality as something necessary and universal. This does not require ontological independence of the faculty of reason from nature or from beings that possess that faculty.

Transcendental arguments show the *necessary* conditions for organisms that might evolve an ability to represent an objective experience. For example, the transcendental necessity of representing nature as consisting of causal relations is an independent condition of a possible organism representing an objective experience, rather than being dependent on nature itself having causal relations. The fact that nature consists of causal relations does not ground the subject’s *a priori* representation of causality. Rather, *a priori* elements of our cognition, such as space, time, substance, and causality, make it possible for us to represent the world in terms of causes and effects. Similarly, the transcendental necessity of representing an objective and categorically normative morality is not dependent on the actual evolution of these cognitive structures, which is an empirical matter. Transcendental necessity, on the other hand, is not an empirical matter, just as logical implications. This means that the transcendental justification of pure reason (*Wille*) is independent of particular organisms, but the *existence* of that faculty depends

on particular organisms possessing a mind with a certain level of complexity. There is no need, on this picture, to postulate a separately existing pure reason floating free of particular organisms that possess it. Thus, Kant's transcendental method does not necessarily have ontological implications.

How is this picture related to the kind of ontology described above? According to Parfit's ontology, even if nothing had ever existed, there would have been moral principles, which exist in a non-ontological sense. Parfit sees moral principles, along with mathematical and logical principles, as independent of particular organisms: morality is stance-independent. As opposed to this realist picture, transcendental constitutivism claims that pure reason (understood as freedom or *Wille*) is the metaphysical ground of the moral law (the CI); however, the existence of pure reason depends on the existence of beings that possess that faculty. This means that if nothing had ever existed, the moral law would not have existed.⁸⁸ In this sense, morality is *stance-dependent*: it is dependent on a particular human (or rational) stance. The existence of morality, in other words, depends on the existence of rational and free beings. Even if the transcendental necessity of representing an objective and categorically normative morality is still independent of the existence of rational and free beings, that does not mean that morality is stance-independent. The necessary relation (the transcendental necessity) between freedom and morality could be independent of the existence of free beings: the fact that "*if* there are free beings, they are necessarily under the moral law" is independent of whether there are free beings. But this would merely be a methodological justification of morality rather than its metaphysical source.

⁸⁸ Or it would perhaps exist only in a wide sense, just as things that are merely possible.

Second, Kant's conception of autonomy does not require metaphysically queer entities. I argued in 3.3.4 that Humean constitutivism does not endorse the *causal independence* aspect of Kant's autonomy. So, Humean constitutivism cannot capture the moral phenomenon of having a bad conscience, the thought that you acted according to your strongest desire but at the same time your action was morally wrong (cf. 3.1).⁸⁹ Transcendental constitutivism can capture causal independence without having to commit itself to metaphysically queer entities. The distinction between *causal (structural) independence* and *existential (transcendent) independence* explains this idea very well: just as a wax cannot change the mold when they are in contact, so the content of the law prescribed by reason remains unaffected by natural causation, even though our reason resides in nature. The following is how Rauscher explains causal or structural independence:

A can be "independent from determination" by B [in] that A may exist in contact with B but consist of such a structure that B is unable to alter A. This is structural independence. A clear example lies in a sealing wax mold used on hot wax applied to an old parchment letter. The mold itself has contact with the wax while remaining unchanged by the wax. Because of the relative malleability of the wax and rigidity of the mold, on the one hand, the wax is not capable of changing the mold at all. The mold, on the other hand, is capable of determining the shape of the wax. [...] This is the idea behind the structural freedom of reason: reason is said to exist in nature but as an unalterable structure of thinking that then processes empirical inputs in a manner independent of – and unchanged by – that empirical input (Rauscher 2015, 126).

Structural dependence contrasts with existential or transcendent dependence, according to which things exist apart from each other in a way that they lack contact. For example, if I am in quarantine with COVID-19, I lack physical contact with other people. This is similar to the idea that pure reason exists as a thing in itself in a noumenal realm

⁸⁹ This phenomenon reflects the categorical (desire-independent) nature of morality.

such that space and time do not contact pure reason. Transcendental constitutivism does not adopt this idea.

Pure reason, according to this interpretation, is a transcendently identifiable, emergent function intrinsic in empirical reason in nature. In other words, reason exists in nature as an unchangeable structure of cognition and thinking. On this picture, pure reason originates from natural causes. But once our mind reaches a certain level of complexity, that is, once we meet the conditions of reason, our reason acquires a function such that it processes empirical input in a way that it is unchanged by that input. This emergent function not only generates a priori cognitions, concepts, ideas, and judgments that make it possible for us to represent the physical world in a certain way; it also generates the moral principle (the moral law) that provides the formal systematic foundation for practical decisions. Granted, any particular instantiation of the conditions of reason is a contingent result of natural causation, but the validity of the moral law is defended through the transcendental argument which asserts that the CI arises from pure reason rather than from contingent foundations in nature such as the evolutionary history of organisms. That is, no matter how exactly organisms have come to possess reason, the content of what reason prescribes (or the way reason functions) never changes (KrV B583–4).⁹⁰ Just as water starts boiling once a certain threshold (100°C) has been reached, so reason spontaneously comes up with the necessary and universal moral law when organisms such as humans meet the conditions of reason and start thinking about what to

⁹⁰ We can liken the structure of reason to the laws of logic: it does not change, and it does not have any specific duration.

do.⁹¹ Without a doubt, a brain damage or a drug can change how one's reason functions, but the content of what reason demands does not change under any circumstances.

The fact that transcendental constitutivism can be given a naturalistic interpretation could give rise to a new question: if reason resides in nature, how could it generate a categorical foundation for morality? How could any natural fact or property provide desire-independent reasons for action (cf. 1.4.2)? However, this does not necessarily pose a problem for transcendental constitutivism because it does not reduce moral value to *any* property, natural or non-natural. On this view, value is conceived of as an *operating principle* of how our reason functions rather than as an *entity* or *property* that can be detected under an ontological radar. Even though pure reason arises from natural causes and the emergent function or the structure of reason could be seen as existing in nature, the structure of reason is not affected by natural causes in the sense that the *relation* between this structure and the moral law exists in a non-ontological sense, just as logical and mathematical principles. For example, logical laws are not properties, nor the validity of them is defended through properties that exist in the actualist sense. Rather, they are constitutive of how we engage in theoretical reasoning: we have no option but to think in terms of logical laws once our physical nature reaches a certain level of sophistication.⁹² Likewise, the CI does not involve (and is not justified by) any value property, but rather it governs our thinking about practical matters when we cross an evolutionary threshold.

⁹¹ The necessary connection between the conditions of reason and morality has been emphasized by some evolutionary biologists as well: "The necessary conditions for ethical behavior [...] depend on an advanced intelligence [...] and] only come about after the crossing of an evolutionary threshold" (Ayala 2016, 250).

⁹² "The laws of logic govern our thoughts because if we don't follow them we just aren't thinking" (Korsgaard 2009, 32).

Just as logical and mathematical principles do not depend on our desires and inclinations, so our reason functions in a way that it is not affected by our desires and inclinations: it functions in a *categorical* way.

Transcendental constitutivism focuses on how our reason must function if we are to represent an objective and categorically necessary morality. One might claim that the way the transcendental constitutivist describes the function of our reason is at odds with the basic constitutivist strategy to ground moral truths in a less controversial descriptive foundation. So, a further objection could be that even if there is no appeal to a noumenal realm, this is still a non-trivial view about how our reason functions. The worry could be that transcendental constitutivism gives us a non-trivial conception of the nature of rationality because Kant's conception of rationality, on this view, involves (1) a commitment to the CI, and (2) the concept of transcendental freedom for which no proof can be given.

First, many philosophers regard the formulations of the CI as implausible, which is an indication that Kant's conception of rationality is non-trivial. People who reject Kant's moral theory or his transcendental idealism would not regard transcendental constitutivism as a plausible metaethical theory. Second, Kant seems to deny that we can give a transcendental proof of the necessity of the moral law (the CI) for our moral experience. The transcendental deduction, according to Kant, proves that a priori elements of cognition are necessary for representing an objective experience (KrV B159). In Kant's view, the moral law is not subject to a transcendental deduction unlike the categories of the understanding because it lacks an object in experience. When it comes to the transcendental deduction of the categories of the understanding, there are objects of

experience to be synthesized, and the categories constitute the rules for synthesis. But since the moral law is not associated with possible objects in experience, Kant thinks he cannot provide a deduction in the same way. On the contrary, the moral law grounds the existence of objects in the sense that it constitutes the foundation for practical decisions and brings about actions:

With the deduction, that is, the justification of its [the moral law] objective and universal validity and the discernment of the possibility of such a synthetic proposition a priori, one cannot hope to get on so well as was the case with the principles of the pure theoretical understanding. For, these referred to objects of possible experience, namely appearances, and it could be proved that these appearances could be cognized as objects of experience only by being brought under the categories in accordance with these laws and consequently that all possible experience must conform to these laws. But I cannot not take such a course in the deduction of the moral law. For, the moral law is not concerned with cognition of the constitution of objects that may be given to reason from elsewhere but rather with a cognition insofar as it can itself become the ground of the existence of objects and insofar as reason, by this cognition, has causality in a rational being, that is, pure reason, which can be regarded as a faculty immediately determining the will (KpV 5:46).

Kant's concern here is not that we cannot become aware of the conception of the moral law as a principle arising a priori out of pure reason or freedom. Rather, his concern is that we cannot infer that human beings really are transcendently free and are subject to this a priori principle. Kant defines transcendental freedom as

a causality in our power of choice such that, independently of those natural causes and even opposed to their power and influence, it might produce something determined in the temporal order in accord with empirical laws, and hence begin a series of occurrences entirely from itself (KrV B562f.).

Kant denies that we can give a proof that human beings really possess this kind of freedom and are really subject to the moral law. There is simply no strong basis to prove this. Our cognitive faculties are simply not sufficient to show that we are in fact free in

this sense or why (if we are in fact transcendently free): “all human insight is at an end as soon as we have arrived at basic powers” (KpV 5:47).

Admittedly, transcendental constitutivism is a non-trivial view about how reason functions. One might deny that the CI is constitutive of rationality, or that our reason functions in the way Kant describes. This, however, does not pose a problem for this dissertation because my claim is not that transcendental constitutivism is the correct metaethical position. Rather, my claim is simply that transcendental constitutivism is a neglected alternative to moral realism. Furthermore, it is possible to find ‘companions in innocence’ that place the *function* of reason at the center of their theories, such as Chomskyan linguistics, Mikhail’s (2007) universal moral grammar, functionalism in the philosophy of mind, and Fodor’s modularity of mind. So, the basic idea behind transcendental constitutivism is not far-fetched.

According to transcendental constitutivism, the moral law is not based on an objective reality that our mind adapts to but rather it is an “a priori proposition that is not based on any intuition, either pure or empirical” (KpV 5:31). This indicates that Kant uses the same conception of knowledge that he developed in the *Critique of Pure Reason*.⁹³ On this conception, there is no a posteriori or a priori knowledge. Our cognition possesses certain elements prior to experience, but these a priori elements do not constitute knowledge. Rather, they are the conditions that makes knowledge possible. If we are to have objective and universal knowledge of the world, our cognition must have a priori elements, such as space, time, substance, and causality, which constitute the conditions of the possibility of knowledge. In other words, our mind must know how to arrange our

⁹³ Cf. Sensen 2017.

experiences before we have any experience. Similarly, if morality is to be objective and universal, it must lie a priori in our cognition, since “necessity and strict universality are [...] secure indications of an *a priori* cognition” (KrV B4). We can imagine the moral law as an innate principle that governs our practical decisions before we make any decisions, even though it is not strictly innate (3.4.4.2).⁹⁴ Our reason functions in a way that organizes our moral experience before we have it and governs our decisions before we make them. Noam Chomsky’s linguistics employs this concept of a priori function.

Chomsky (1957, 1965) made a Kantian revolution in linguistics. He revolutionized linguistics such that he changed the main question of linguistics from “What are the differences in different languages?” to “What are the common features shared by all languages?” According to Chomsky’s theory of universal grammar, there is more to language acquisition than mere imitation. First, some animals can imitate certain expressions but that does not amount to learning a language or mastering the grammar. Second, children use words that they never heard, such as ‘broke,’ ‘bought,’ or ‘best.’ Chomsky’s solution employs Kant’s idea that “objects must conform to our cognition” (KrV Bxvi). (In Chomsky’s case, of course, it is language that conforms to our cognition.) He asserts that there is an a priori aspect of language that organizes our linguistic experience and governs the acquisition of language. Chomsky calls this a priori element of our mind ‘Universal Grammar,’ and as the name suggests, it is shared by all human beings. Universal Grammar explains why human children have the ability to learn any human language. More importantly, it indicates that human languages are grounded

⁹⁴ Kant’s rejection of innate principles seems to be what distinguishes transcendental constitutivism from theories such as Chomsky’s theory of Universal Grammar, Mikhail’s theory of Universal Moral Grammar, and Fodor’s modularity of mind. More on this in 3.5.2.

in certain basic principles that are constitutive of how our reason functions. This idea is consistent with transcendental constitutivism: just as certain basic principles, which are constitutive of how our reason operates, are common to all human languages, so certain basic principles, which are constitutive of how our reason operates, are common to all different moral systems. Since animals do not meet the ‘conditions of reason,’ their cognitive faculties do not function like ours, and thus they cannot learn human languages. They are also not subject to moral principles.

John Mikhail’s (2007) theory of Universal Moral Grammar follows Chomsky’s lead. On his view, each human being possesses intuitive, unconscious, and a priori rule system that constitutes the foundation of moral knowledge and governs our moral judgments. Universal Moral Grammar (or innate moral knowledge) is constitutive of how our reason functions, and it is “a complex and [...] domain-specific set of rules, concepts, and principles” that organize our moral experience and determine intuitive moral judgments (Mikhail 2007, 144). This a priori system provides the framework of possibilities and enables us to determine the moral status of “an infinite variety of acts and omissions” (ibid.). In other words, these basic rules and principles are constitutive of the inherent structure of the mind, but their ontogenetic development must be shaped by one’s environment. This allows for differential moral judgments, but the fact that (1) certain prohibitions such as that on murder, rape, and other types of aggression seem to be universal or nearly so, and that (2) moral dilemmas elicit rapid, intuitive, and widely shared judgments that are made with a high degree of certainty indicate the possibility of an a priori pattern of organization imposed on the given information or stimulus by our mind to produce moral judgments.

According to Mikhail, Universal Moral Grammar (UMG) is an a priori function of our mind which plays a crucial role in interpreting particular cases and assigning them a moral status. Similar to the categories of the understanding, then, the function of UMG is to organize our moral experience and guide our moral judgments. UMG has three elements: (i) deontic rules, (ii) structural descriptions, and (iii) conversion rules. Conversion rules convert the stimulus into an appropriate structural description. This means that we can compute a full structural description of an action by attributing them properties such as ends, means, side effects, and prima facie wrongs, even when actions contain no direct evidence for these properties (2007, 146). Conversion rules “generate a complex representation of the action that encodes pertinent information about its temporal, causal, moral, intentional, and deontic properties” (2007, 148). They compute the temporal order, underlying causative and semantic structures, and intentional structures of actions and apply certain moral principles to these structures to produce representations of good and bad effects.⁹⁵ This indicates that when people are presented with moral dilemmas, they unconsciously compute structural descriptions of them. People can assign a moral status (e.g., obligatory, permissible, forbidden) to these cases because they seem to have a tacit knowledge of legal or moral rules, such as the wrongness of intentional battery and the principle of double effect.⁹⁶

⁹⁵ For example, people’s reactions to the original trolley case indicate that they make an a priori distinction between ‘battery as a means’ and ‘battery as a side effect.’ Moreover, on Mikhail’s view, the context in which a dilemma is presented affects the nature of our structural description of it and the way we apply the conversion rules. This seems to confirm my claim in 2.2.3 that the representative context of the ‘Bus dilemma’ enables most of us to make a moral distinction between ‘harming an innocent person’ and ‘directing a public threat to a lesser harm.’

⁹⁶ According to the principle of double effect, (a) if an action has both good and bad consequences, (b) if the bad consequences are not directly intended, and (c) if the good consequence outweighs the bad one, then the act is permissible.

Individuals are intuitive lawyers who possess a natural readiness to compute mental representations of human acts in legally [and morally] cognizable terms [... and] who implicitly recognize the relevance of ends, means, side effects and prima facie wrongs, such as battery, to the analysis of legal and moral problems (2007, 145–6).

Even though it far from obvious that Mikhail's account is correct, it has considerable explanatory benefits. First, it seems to explain why every language has phrases to express basic moral concepts, such as 'obligatory,' 'permissible,' 'forbidden,' and so on. Second, it seems to explain why even a very young child can make a distinction between moral rules and mere conventions (cf. Nichols 2004). Third, it seems to explain why condemnation of murder, rape, and other types of intentional battery, as well as moral distinctions based on causation and intention, are universal or nearly so, even though some cultures may have further rules (e.g., honor codes) that justify these condemnable actions more easily than some other cultures. Fourth, it seems to explain why some moral dilemmas elicit rapid, intuitive, and widely shared judgments that are made with a high degree of certainty. And fifth, since UMG is an unconscious and a priori rule system, it seems to explain why people often have difficulty giving a compelling justification for their moral intuitions.

Lastly, Chomsky's and Mikhail's theories are not the only ones that place a strong emphasis on the function of reason. Functionalism in the philosophy of mind and Jerry's Fodor's modularity of mind can also be given as examples. Functionalism claims that mental states are not defined by their internal constitution but by the *role* they play in the system of which they are a part. The conception of the mind as a system of functions dates back to Kant, who thinks of the mind as a system of conceptual functions that transform objects of perception into representations. Moreover, the idea that mind is made up of independent special-purpose modules also dates back to Kant, who introduces

the concept of modular organization in his theory of mental faculties. This idea paved the way for Fodor's (1983) famous modularity account in his book *The Modularity of Mind*. According to Fodor, our reason functions in a way that transforms the unprocessed sense data into formats each innately specified module can process. These modules are hardwired, fast, autonomous, stimulus-driven, and insensitive to central cognitive goals. This account bears similarities to Kant's description of the categories of the understanding and the transcendental constitutivist interpretation of Kant, according to which the moral law is an a priori principle that organizes our moral experience and governs our practical decisions. A priori elements of cognition (e.g., space, time, substance, causality, the moral law) can be thought of as hardwired, fast, autonomous, stimulus-driven modules that constitute the framework for our perceptual and moral experience and the conditions of the possibility of perceptual and moral knowledge.

3.5.2 The Bootstrapping Objection

The idea that the moral law, namely the CI, is constitutive of how our reason functions invites another question: how can we bootstrap substantive reasons into existence from a merely formal moral law? In other words, how can we derive substantive morality out of thin air? This objection is motivated by Hegel's (1991, §135) famous 'empty formalism' charge against Kant, according to which Kant's moral theory lacks content, authority, and motivation because it is based on a purely formal principle. If we resort to a minimal conception of reason that is purely formal, how can we vindicate anything substantive? This objection, I believe, is an important factor that pushes some moral philosophers into Humean constitutivism.

The relevant question here is, how are we supposed to derive specific moral content from the CI? How does Kant's universalizability principle generate particular moral duties? The traditional understanding of such derivation involves three steps: (1) you must identify your maxim, that is, your subjective principle or reason to perform a certain action (e.g., I will lie in order to get what I want); (2) you must raise your maxim to the level of a universal law such that your subjective principle operates with the same regularity of the laws of nature (e.g., Everyone will lie whenever they can get what they want); (3) you must determine whether your maxim, when raised to the level of universality, leads to a contradiction. If it leads to a contradiction, then you know that the action is prohibited.⁹⁷

The type of contradiction that determines the wrongness of an action can be interpreted in different ways. It could be that universalizing the maxim "I will lie whenever I can get away with it" will lead to a *logical* contradiction: the only way for me to get away with my lie is on the presumption of truth-telling. Others must assume that I will tell them the truth in order for me to successfully lie to people. But if there is no such assumption when the maxim is universalized, it is impossible for me to lie. Or if everyone steals all the time, there will be no private property, and this will render stealing impossible because stealing presupposes private property. Universalizing such maxims would be as impossible as conceiving four-sided triangles. Another way of thinking about such a contradiction is to view it as a *self-contradiction* of the will,⁹⁸ which occurs when

⁹⁷ For the traditional understanding of the derivation of specific duties from the CI, see also Rawls 2000, 167–70.

⁹⁸ Kleingeld 2017.

your maxim and the universalized version of it cannot be willed at the same time: you want to lie to get what you want, *and* you want everyone to lie to get what they want. You do *not* want to lie to get what you want when you universalize your maxim because you realize that you cannot get what you want under those circumstances, but that contradicts your initial attitude: you wanted to lie to get what you want. Third way of thinking about such a contradiction is to view as a *practical* contradiction.⁹⁹ If people know that others will always tell a lie to get what they want with the same regularity of the laws of nature, then nobody will believe what I say. But that means I cannot achieve my original purpose of getting what I want: I defeat my own purpose.

The problem is that it is not clear why the above types of contradictions should matter *morally*. First, why would a logical contradiction amount to a moral problem? Why would a four-sided triangle, for example, be morally problematic? Second, self-contradiction of the will occurs quite often, as we often want to do something and do not want to do that very thing at the same time: I want to meet my ex-girlfriend and I don't want to meet her at the same time, but I am not sure why this contradiction itself would lead to a moral problem. Third, it is not clear why not being able to get what you want would be morally relevant. If I want to meet my ex-girlfriend and do something that defeats my purpose, I am perhaps imprudent or irrational but not necessarily immoral. A serial killer could skillfully implement his plan of killing his victim and get what he wants, but he is still immoral.

Kant famously says, "without sensibility no object would be given to us, and without understanding none would be thought. Thoughts without content are empty, intuitions

⁹⁹ This is Korsgaard's (1996, 92–101) interpretation.

without concepts are blind” (KrV B75). This means that a priori elements of cognition cannot produce anything by themselves. They cannot form knowledge by themselves. For us to perceive anything or have knowledge, there must be objects of possible experience or appearances, so that they can be brought under the categories of the understanding. We have seen that transcendental constitutivism conceives of the moral law as one of the a priori elements of cognition, so we can apply the same idea to morality: a purely formal law cannot produce anything by itself; empirical matter must be brought under the moral law to generate specific moral content. Kant seems to agree: the a priori moral law “needs anthropology for its *application* to human beings.” (GMS 4:412) Kant calls the empirical part of morality “practical anthropology” and he calls the rational part of it “morals” (GMS 4:388).

Kant also says, “there must [...] be rules whereby my actions hold good universally, and these are derived from the *universal ends of mankind*, and by them our actions must agree; and these are moral rules” (VE 27:258; emphasis added). When deriving specific moral duties, we want that “a certain principle [to] be objectively necessary as a universal law” (GMS 4:424). We want universal laws because laws must not serve our private, self-interested ends if they are to be morally relevant. These laws, which we regard as objectively necessary in this sense, could be revealed by empirical sciences. Empirical sciences can discover universal human ends and tendencies, such as promoting self-preservation and happiness, avoiding pain, cultivation of social relationships, cultivation of rational capacities, and so on. We can then apply the CI to these universal ends that we regard as objectively necessary: “we shall often have to take as our object the particular *nature* of man, which is known only by experience, in order to *show* in it what can be

inferred from universal moral principles” (MS 6:217). So, (1) we need empirical data to derive specific moral content from the formal CI, and (2) this content must be derived from the universal ends of human beings. This seems to be how we can derive moral duties for human beings.¹⁰⁰

What kind of contradiction is involved in the CI? It seems that Kant associates contradiction with making an exception for yourself to a law you regard as objectively necessary.¹⁰¹

Consequently, if we weighed all cases from one and the same point of view, namely that of reason, we would find a *contradiction* in our own will, namely that a certain principle be objectively necessary as a universal law and yet subjectively not hold universally but allow *exceptions* (GMS 4:424; emphasis added).

This kind of exception occurs when you want other people to follow a rule that you believe to be objectively necessary but you at the same time do not want this rule to apply to you. Admittedly, making an exception for yourself, by itself, is not morally relevant. For example, a student could study more than her classmates to be the best in class, but that kind of exception is not immoral. Or, in a world where everyone cheats on their partners, people who are thinking about cheating on their partners would not be making an exception for themselves. The kind of exception that is morally wrong is when one acts against the maxim “of not exalting oneself above others” (MS 6:449). In other words, if you think that an objective rule should not apply to you simply because you are

¹⁰⁰ Sensen (2014, 2022) and Herman (1993) interpret Kant’s universalizability principle in this way, even though they differ on the nature of universal human ends. While Herman associates the “true needs” of human beings with the ends that *rational*, end-setting agents have (1993, 55), on Sensen’s interpretation, universal human ends are not limited to the ones that support rationality (2022, 5). I believe that this interpretation could help transcendental constitutivism overcome the bootstrapping objection, regardless of whether Kant himself adopts this view.

¹⁰¹ Sensen 2014, 172.

superior to others, then you are making an exception for yourself in a morally problematic way. This is morally problematic because you simply do not regard yourself as one among equals and you think you deserve more than others simply because you are you. The student who tries to be the best by studying hard does not necessarily see herself as more important than others simply because she is herself. But it is morally wrong when one cuts in line at the movies simply because she thinks she is superior to others. If we interpret the CI in this way, then the principle becomes noticeably moral. The form of exception prohibited by the CI reflects the notion of fairness. Unequal treatment on the basis of the fact that ‘one is oneself’ is not morally permissible.

According to the above interpretation, (1) we must consult empirical sciences to discover universal and objective laws that reflect *universal human ends*; (2) we must determine (mid-level) *principles* that are *necessary* for promoting these ends; (3) we must determine whether one is making an *exception* for themselves to the principles they regard as objectively necessary. Universal human ends do not yield necessity by themselves because they are the results of empirical investigations. It is only when our reason makes an inference from these universal ends that necessity enters the picture: “reason alone is capable of discerning the connection of means with their purposes” (KpV 5:58f.). For example, nutrition is a necessary means to the universal end of survival. But still, the necessity here is not yet a *categorical* one because universal ends are contingent upon our nature, which is subject to change. This means necessary means to these ends are also subject to change. Necessary means are *objective* in the sense that they do not depend on personal preferences or private ends, although different cultures might have different rules for, for example, how people should protect or preserve

themselves.¹⁰² But these necessary means and the rules associated with them are not morally relevant because they are not categorically normative: “an end that every man has (by virtue of the impulses of his nature) [...] can never [...] be regarded as a duty” (MS 6:386). These necessary means or rules become morally relevant when our reason requires us not to make an *exception* to these rules, in the sense of not seeing ourselves as superior to others. And these rules become binding through the CI because our maxims must be universalizable: “the only way this maxim [of self-love] can be binding is through its qualification as a universal law” (MS 6:393). According to transcendental constitutivism, the formal, a priori requirement of “not exalting oneself above others” provides the categorical aspect of morality.

A possible objection to this account is that morality amounts to nothing more than a set of prudential guidelines and thus it can only be hypothetically, as opposed to categorically, necessary. On this account, we must take necessary means to the universal ends such as survival, cultivating social relationships, or developing our talents, but, as the objection goes, we are obliged to do so *only if we want* to survive, cultivate social relationships, or develop our talents. However, universal human ends are not the only material from which we derive specific moral content. The fact that all human beings want to survive and reproduce does not mean we are *morally* required to take the necessary means to these universal ends: “what everyone already wants unavoidably, of

¹⁰² Complying with some principles or commands, such as, “Protect and preserve yourself,” or “Look after your children,” are necessary means to survival, but these principles could be interpreted in different ways in different cultures, and thus mid-level principles could vary from culture to culture. For example, different cultures might adopt different rules against murder, or they might adopt different strategies to look after their children. Traffic rules serve the end of protecting lives, but different societies have different drink-driving laws, different seat-belt laws, and different speed limits. So, adoption of these mid-level principles is partly a cultural matter. This seems to be the reason there always has been and will be moral disagreement.

his own accord, does not come under the concept of *duty*” (MS 6:386). However, discovering universal human ends is only the first step of deriving moral content. We must also figure out whether one is making an exception (in the sense of exalting oneself above others) to the laws they regard as objectively necessary. The derivation of specific moral content is governed by the formal, a priori restriction imposed by our reason via the CI.

The objection that no substantive content can be derived from Kant’s CI is hasty. Without a doubt, a purely formal principle, devoid of empirical content, considered in and of itself, is empty:

the [moral] law determines the mode and manner in which a person is obligated [...] and thus abstracts entirely from the actions which he must consequently perform. If, by this, and by difference of obligation, we now divide morality, as theory of conduct, *in genere*, no rule of dutiful action can then itself be determined, *because this belongs to the matter* (VE 27:578; emphasis added).

Nevertheless, if our reason draws inferences from *empirically identifiable* universal human ends to determine specific moral content and makes the necessary means to those universal ends binding through the CI, the objection could fail.

3.5.3 Evolutionary Challenge Revisited

Recall Joyce’s claim that our sense of categorical normativity and our capacity for normative guidance are innate. That is, on Joyce’s view, our belief in categorically normative requirements (our moral sense) is simply encoded in our genes to make us better social cooperators. It could be that the sense of categorical normativity, i.e., the feeling that performing or avoiding certain actions are absolute and unconditional, has been acquired over the course of human evolutionary history and developed or learned

during ontogenetic development. In other words, the essential features of morality, namely objectivity and categorical normativity, could just have been historically conditioned. According to Kant, moral value consists not just in doing the right thing but also in having the proper motivation, namely doing the right thing simply because it is the right thing to do (GMS 4:390). According to Joyce, both aspects of morality (the content and the proper motivation) have been developed to increase our chances of survival and reproduction.

The possibility that our sense of moral objectivity and categorical normativity is merely the product of human evolution threatens to undermine the reliability of our moral beliefs and intuitions and the power of the Normativity Objection (1.4.2). This is because we have a plausible scientific explanation for why we have this moral sense, and non-scientific explanations could be superfluous or implausible. If we have the sense of categorical ought simply because of our evolutionary history, it could be an enormous coincidence if morality turned out to be exactly as expected, i.e., if our evolutionarily shaped moral sense got gene-independent moral truths right. In the following, I will claim that transcendental constitutivism is perfectly compatible with the idea that our moral sense has increased our chances of survival and reproductive success.

In Kant's view, morality "is to manifest its purity as sustainer of its own laws, not as herald of laws that an implanted sense" (GMS 4:425). So, he seems to be against evolutionary explanations of morality. This is because, if it were true that we have internalized a moral sense during our evolutionary history, (or if it were true that a moral sense has been implanted in us by a divine being), then morality would lack necessity. We could have had a different moral sense under different circumstances: "experience

teaches us, to be sure, that something is constituted thus and so, but not that it could not be otherwise” (KrV B3). As I described in 3.4.4.2, Kant denies that a priori elements of cognition have been implanted in us by a divine being (KrV B167) or by natural processes (GMS 4:425). Kant thinks that for something to be universal and necessary it must arise a priori from our reason (KrV B4). Thus, both the categories of the understanding and the moral law are a priori contributions of our reason: they are “originally acquired” (ÜE 8:222). Categories do not exist prior to objects of possible experience or appearance. Rather, once one is confronted with unprocessed appearance, the understanding spontaneously provides the categories, such as space, time, substance, and causality, out of itself and thereby shapes our sense experience (ÜE 8:221). Similarly, once we start deliberating about what to do, about reasons for our actions, our reason “with complete spontaneity [...] makes its own order according to ideas [...] it even declares actions to be necessary” (KrV B576). Just as we perceive the objects and events of the world in terms of cause and effect, so our practical deliberation is governed by the moral law, which arises a priori from our reason. The moral law is therefore the product of a spontaneous activity of reason.

The important point here is that our cognitive faculties that generate the categories of the understanding or the moral law (or pure reason in general) could be innate in the sense of being the result of evolutionary forces, but the *generation* of these a priori cognitions, concepts, and principles by our reason is a *spontaneous, necessary, and history-independent* activity of our reason. As I described in 3.5.1, pure reason can be seen as existing in nature as a transcendently identifiable, emergent function or structure inherent in empirical reason. In other words, our mind reaches a certain level of

complexity due to evolutionary processes, and once our mind crosses a certain threshold, our reason acquires a function such that it processes empirical input in a way that it is unchanged by that input. (This is the sense in which rational beings are free: they can act against their strongest desires.) This emergent structure generates the categories of the understanding and the moral law, thereby providing unity and systematic connection among theoretical and practical concepts, but the structure itself can be seen as existing in nature. So, it is possible that this structure has been acquired over the course of our evolutionary history. However, neither the categories nor the moral law have been *evolved* from the structure of reason itself (pure reason or freedom). Rather, they follow from this structure *necessarily*. This means if we could create a free being whose mind is structurally independent of natural causes, i.e., a being with a mind that has a certain level of complexity, *out of nothing*, that being would be under the moral law and would have a sense of moral ought. This seems to be the sense in which our sense of categorical ought is *not* innate, and the CI is a timeless principle. And this seems to be the sense in which reason “does not alter” (KrV B584). Evolutionary history is not required to explain how the CI is generated.

Are we really free in the sense that we can act against our strongest desires? Kant thinks we cannot know whether and why we are free: “the moral law is, in fact, a law of causality through freedom ... [However, it] cannot be proved by any deduction ... to answer this surpasses every faculty of our reason” (KpV 5:47; KrV B585). As seen above, Kant denies that we can provide a transcendental deduction of the moral law because the moral law lacks an object in experience (KpV 5:46). Although the moral law has no objects of experience for which it can be a transcendental foundation, Kant

believes that there is an *experience* through which we can discover the moral law: the experience of a rational being who deliberates on what to do and realizes that she can act against her strongest desire. That is, we sometimes feel the *necessity* of doing the right thing despite our strongest desires and inclinations, and, Kant thinks, this sense of necessity gives us a justification for believing that we are free: “we can become aware of pure practical laws just as we are aware of pure theoretical principles, by attending to the *necessity* with which reason prescribes them to us and to the setting aside of all empirical conditions to which reason directs us” (KpV 5:30; emphasis added).

Kant gives an example where a person, who knows that he is under a moral requirement to tell the truth, is going to be severely punished (perhaps along with his family) if he refuses to give false testimony:

Ask him whether, if his prince demanded, on pain of the same immediate execution, that he give false testimony against an honorable man whom the prince would like to destroy under a plausible pretext, he would consider it possible to overcome his love of life, however great it may be. He would perhaps not venture to assert whether he would do it or not, but he must admit without hesitation that it would be possible for him. He judges, therefore, that he can do something because he is aware that he ought to do it and cognizes freedom within him, which, without the moral law, would have remained unknown to him (KpV 5:30).

The point of the example is to show that we are justified in believing that we are free, in the sense that we are able to start a series of event in nature without being caused by the forces of nature. The person in Kant’s example becomes aware of the necessity and the obligation of not giving false testimony, and this awareness implies that he *can* act against his strongest desires: ought implies can. He realizes that he is free through becoming aware of the possibility that he can decline giving false testimony. That is, our subjective awareness of the moral law and the accompanying feeling of obligation and necessity let us discover our freedom: “the moral law is the *ratio cognoscendi* [epistemic

ground] of freedom” (KpV 5:4). We would not be able to know that we are free without the experience of the categorical ought that claim authority over our desires and inclinations.

Kant then attempts to justify the existence of the CI from the existence (in fact, the assumption) of freedom, which is the “*ratio essendi* [metaphysical ground] of the moral law” (ibid.). Kant regards freedom as a form of causality. That is, freedom can bring about change in the world. Since “the concept of causality brings with it that of laws,” freedom must have a law (GMS 4:446). This law, according to Kant, can only be the CI, as shown by the above example: the person who is contemplating about whether to give false testimony realizes that he has freedom or causality that is independent of any of his desires. While freedom is the metaphysical source of the CI, the CI enables us to discover our freedom. It is possible, on this account, that our freedom has been evolved. But the moral law has not evolved from freedom. Rather, the moral law follows necessarily from freedom.

A disadvantage of the transcendental constitutivist reading is that it seems to be unable to explain *why* we should follow the CI. Humean constitutivism and evolutionary accounts seem to do better on this front because they can give us independent reason for the usefulness of morality. If following moral principles enhances our survival prospects, then we have good reason to follow these principles. Similarly, if following the dictates of our reason is evolutionarily advantageous or if it enables us to satisfy our desires or preferences, then we should take an interest in doing what our reason says. But why should we be interested in complying with the categorical requirements of morality when doing so will prevent us from achieving something extremely desirable (or avoiding

something extremely painful)? Kant's answer is that our desires and inclinations are part of our animal nature, whereas the demands of our reason are part of our rational nature, and thus moral obligations must have priority over desires: "the law interests because it is valid for us as human beings, since it arose from our will as intelligence and so from our proper self" (GMS 4:461). This answer appears to be less satisfactory than a Humean or an evolutionary answer to the question of moral motivation. It may be difficult to explain why we should be interested in an unconditional law given spontaneously by our reason. But it may also be impossible to provide a fully convincing answer to question 'Why be moral?' because we can never really eliminate the possibility of indifference. If the demands of our reason carry categorical normative force, the possibility of an agent who refuses to care about moral requirements does not really undermine the reason-giving force of morality. But in any case, even if we concede that Humean constitutivism is better equipped to explain our interest in morality, we must also accept that Humean accounts destroy the essential features of common sense morality, namely objectivity and categorical normativity. Transcendental constitutivism, on the other hand, can capture those features and it is a neglected alternative to moral realism, which is all I need to show.

In closing this dissertation, I would like to mention further potential disadvantages of transcendental constitutivism. First, transcendental constitutivism depends on two questionable assumptions: (1) we are free in the sense of having the ability to start a series of events in nature without being affected by natural forces, and (2) freedom in this descriptive sense generates a law that reflects the idea of fairness, which is a moral concept. We cannot prove that we have the power to act against our strongest desires. We

cannot also prove that pure reason exists in nature as an emergent, unalterable structure of thinking. On top of that, the bootstrapping or emptiness objection might still pose a problem because it is unclear why the CI, which is the first *normative* reality, would follow necessarily from freedom in a *descriptive* sense. In other words, why would our reason command us to be fair (in the sense of not making an exception to the laws we regard as objectively necessary), even if we have this power to act against our desires? Making these extra assumptions seems to be the main disadvantage of such a view. But as the first chapter and the following discussion show, moral realism suffers not only from making problematic assumptions about moral ontology and moral epistemology but also from alienating us from morality. So, the fact that transcendental constitutivism might have certain disadvantages does not undermine my claim that it is a neglected alternative to moral realism.

Second, my rejection of the ontological characterization of moral value and perceptual characterization of moral knowledge appears to depend partly on the claim that drawing an analogy between empirical facts and moral facts may conflate the kind of objectivity possessed by empirical facts with the kind of objectivity possessed by moral facts. I claimed that since empirical facts are about ‘what *is the case*’ and moral facts are about ‘what *ought to be done*,’ it would not be very surprising if it turned out that moral facts and empirical facts possess different kinds of objectivity. This claim seems to work against realism, but it seems that moral objectivity and the objectivity of empirical facts become similar if we accept transcendental constitutivism. According to transcendental constitutivism, it seems that both the empirical sciences and morality have the same sort of objectivity because in both cases a priori elements of cognition make it possible for us

to have objective and universal knowledge.¹⁰³ For example, we can have objective and universal knowledge in physics because the categories of the understanding provide a systematic connection and coherence among theoretical concepts and enable us to perceive the world in a certain way. Similarly, we can have objective and universal moral knowledge because the moral law provides a systematic connection and coherence among practical concepts and governs our practical decisions. On this view, a priori elements of cognition constitute the conditions of the possibility of both perceptual and moral experience, so the source of both moral objectivity and empirical objectivity lies a priori in our cognitive faculties. Indeed, one might say that transcendental constitutivism comes with high costs because of its commitment to transcendental idealism. This would be a legitimate concern. However, this concern would not undermine my claim that transcendental constitutivism is a neglected alternative to moral realism.

¹⁰³ Of course, Kant thinks there is an important distinction between empirical facts and moral facts because “the moral law is not concerned with cognition of the constitution of objects that may be given to reason from elsewhere but rather with a cognition insofar as it can itself become the ground of the existence of objects” (KpV 5:46). This captures the distinction between ‘is’ and ‘ought.’ However, transcendental constitutivism claims that Kant employs his transcendental idealist insight both in his theoretical and practical philosophy. This seems to indicate that the source of both moral objectivity and empirical objectivity lies a priori in our cognitive faculties.

References

- Allison, Henry E. 1990. *Kant's Theory of Freedom*. New York: Cambridge University Press.
- Arrhenius, Gustaf. 2000. "An Impossibility Theorem for Welfarist Axiologies." *Economics and Philosophy* 16 (2): 247–66.
- Audi, Robert. 2004. *The Good in the Right*. Princeton: Princeton University Press.
- Ayala, Francisco J. 2016. *Evolution, Explanation, Ethics and Aesthetics: Towards a Philosophy of Biology*. Cambridge, MA: Academic Press.
- Bagnoli, Carla. 2013. "Introduction." In *Constructivism in Ethics*, edited by Carla Bagnoli, 1–21. Cambridge: Cambridge University Press.
- Bedke, Matthew. 2009. "Intuitive Non-Naturalism Meets Cosmic Coincidence." *Pacific Philosophical Quarterly* 90 (2): 188–209.
- . 2012. "Against Normative Naturalism." *Australasian Journal of Philosophy* 90 (1): 111–29.
- Benovsky, Jiri. 2015. "Dual-Aspect Monism." *Philosophical Investigations* 39 (4): 335–52.
- Blackburn, Simon. 1984. *Spreading the Word*. New York: Oxford University Press.
- . 1998. *Ruling Passions*. Oxford: Oxford University Press.
- Bojanowski, Jochen. 2015. "Kant on Human Dignity." *Kant-Studien* 106 (1): 78–87.
- Boyd, Richard. 1988. "How to Be a Moral Realist." In *Essays on Moral Realism*, edited by Geoffrey Sayre-McCord, 181–228. New York: Cornell University Press.
- Bratman, Michael E. 2012. "Constructivism, Agency, and the Problem of Alignment." In *Constructivism in Practical Philosophy*, edited by Jimmy Lenman and Yonatan Shemmer, 81–98. Oxford: Oxford University Press.
- . 2018. *Planning, Time, and Self-Governance*. Oxford: Oxford University Press.

- Brink, David. 1986. "Externalist Moral Realism." *Southern Journal of Philosophy* 24 (Supplement): 23–42.
- Bukoski, Michael. 2016. "A Critique of Smith's Constitutivism." *Ethics* 127 (1): 116–46.
- . 2017. "Self-Validation and Internalism in Velleman's Constitutivism." *Philosophical Studies* 174 (11): 2667–86.
- Burkart, Judith M., Rahel K. Brügger, and Carel P. van Schaik. 2018. "Evolutionary Origins of Morality: Insights from Non-Human Primates." *Frontiers in Sociology* 3 (17): 1–12.
- Buss, Sarah, and Andrea Westlund. 2018. "Personal Autonomy." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta. <https://plato.stanford.edu/entries/personal-autonomy/>.
- Cavalli-Sforza, Luigi, and Marcus Feldman. 1981. *Cultural Transmission and Evolution*. Princeton: Princeton University Press.
- Chatterjee, Anjan. 2014. *The Aesthetic Brain: How We Evolved to Desire Beauty and Enjoy Art*. Oxford: Oxford University Press.
- Chomsky, Noam. 1957. *Syntactic Structures*. Berlin: Mouton de Gruyter.
- . 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: The MIT Press.
- Christman, John. 1991. "Autonomy and Personal History." *Canadian Journal of Philosophy* 21 (1): 1–24.
- . 1993. "Defending Historical Autonomy: A Reply to Professor Mele." *Canadian Journal of Philosophy* 23 (2): 281–90.
- Collier, John, and Michael Stiglitz. 1993. "Evolutionary Naturalism and the Objectivity of Morality." *Biology and Philosophy* 8 (1): 47–60.
- Copp, David. 2003. "Why Naturalism?" *Ethical Theory and Moral Practice* 6 (2): 179–200.
- . 2007. *Morality in a Natural World*. Cambridge: Cambridge University Press.
- Cuneo, Terence. 2007. *The Normative Web: An Argument for Moral Realism*. Oxford: Oxford University Press.
- Darwin, Charles. 1981. *The Descent of Man and Selection in Relation to Sex*. Princeton: Princeton University Press.

- Das, Ramon. 2016. "Evolutionary Debunking of Morality: Epistemological or Metaphysical?" *Philosophical Studies* 173 (2): 417–35.
- Dawkins, Richard. 2006. *The Selfish Gene*. Oxford: Oxford University Press.
- Dowell, Janice. 2016. "The Metaethical Insignificance of Moral Twin Earth." In *Oxford Studies in Metaethics 11*, edited by Russ Shafer-Landau, 1–27. Oxford: Oxford University Press.
- Dreier, James. 2004. "The Problem of Creeping Minimalism." *Philosophical Perspectives* 18 (1): 23–44.
- Enoch, David. 2006. "Agency, Shmagency: Why Normativity Won't Come from What Is Constitutive of Action." *Philosophical Review* 115 (2): 169–98.
- . 2009. "Can There Be a Global, Interesting, Coherent, Constructivism About Practical Reason?" *Philosophical Explorations* 12 (3): 319–39.
- . 2011. *Taking Morality Seriously: A Defense of Robust Realism*. Oxford: Oxford University Press.
- . 2014. "Why I Am an Objectivist about Ethics (And Why You Are, Too)." In *The Ethical Life*, edited by Russ Shafer-Landau, 208–21. New York: Oxford University Press.
- Ferrero, Luca. 2010. "Constitutivism and the Inescapability of Agency." In *Oxford Studies in Metaethics 4*, edited by Russ Shafer-Landau, 303–33. Oxford: Oxford University Press.
- . 2018. "Inescapability Revisited." *Manuscrito: Revista Internacional de Filosofia* 41 (4): 113–58.
- . 2019. "The Simple Constitutivist Move." *Philosophical Explorations* 22 (2): 146–62.
- Finlay, Stephen. 2014. *Confusion of Tongues: A Theory of Normative Language*. Oxford: Oxford University Press.
- Firth, Roderick. 1952. "Ethical Absolutism and the Ideal Observer." *Philosophy and Phenomenological Research* 12 (3): 317–45.
- FitzPatrick, William J. 2005. "The Practical Turn in Ethical Theory: Korsgaard's Constructivism, Realism, and the Nature of Normativity." *Ethics* 115 (4): 651–91.

- . 2008. “Robust Ethical Realism, Non-Naturalism, and Normativity.” In *Oxford Studies in Metaethics 3*, edited by Russ Shafer-Landau, 159–205. Oxford: Oxford University Press.
- . 2013. “How Not to Be an Ethical Constructivist: A Critique of Korsgaard’s Neo-Kantian Constructivism.” In *Constructivism in Ethics*, edited by Carla Bagnoli, 41–62. Cambridge: Cambridge University Press.
- . 2014. “Skepticism about Naturalizing Normativity: In Defense of Ethical Nonnaturalism.” *Res Philosophica* 91 (4): 559–88.
- . 2016. “Ontology for an Uncompromising Ethical Realism.” *Topoi* 37 (4): 537–47.
- . 2018. “Representing Ethical Reality: A Guide for Worldly Non-Naturalists.” *Canadian Journal of Philosophy* 48 (3): 548–68.
- Flack, Jessica C., and Frans de Waal. 2000. “Any Animal Whatever: Darwinian Building Blocks of Morality in Monkeys and Apes.” *Journal of Consciousness Studies* 7 (1–2): 1–29.
- Flanagan, Owen J. 1991. *The Science of the Mind*. Cambridge, MA: MIT Press.
- Fodor, Jerry A. 1983. *The Modularity of Mind*. Cambridge, MA: The MIT Press.
- Foot, Philippa. 1967. “The Problem of Abortion and the Doctrine of Double Effect.” *Oxford Review* 1 (5): 5–15.
- . 2001. *Natural Goodness*. Oxford: Clarendon Press.
- Formosa, Paul. 2013. “Is Kant a Moral Constructivist or a Moral Realist?” *European Journal of Philosophy* 21 (2): 170–96.
- Freeman, Dwight, John Graham, and John Emlen. 1993. “Developmental Stability in Plants: Symmetries, Stress and Epigenesis.” *Genetica* 89 (1): 97–119.
- Gibbard, Allan. 1990. *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Cambridge, MA: Harvard University Press.
- . 2003. *Thinking How to Live*. Cambridge, MA: Harvard University Press.
- Gould, Stephen Jay. 1996. *The Mismeasure of Man*. New York: W. W. Norton & Company.
- Gould, Stephen Jay, and E. S. Vrba. 1982. “Exaptation—A Missing Term in the Science of Form.” *Paleontological Society* 8 (1): 4–15.

- Greene, Joshua, R. Brian Sommerville, Leigh E. Nystrom, John M. Darley, and Jonathan D. Cohen. 2001. "An FMRI Investigation of Emotional Engagement in Moral Judgment." *Science* 293 (5537): 2105–8.
- Greene, Joshua. 2008. "The Secret Joke of Kant's Soul." In *Moral Psychology, Vol. 3*, edited by W. Sinnott-Armstrong, 35–79. Cambridge, MA: MIT Press.
- Greene, Joshua, Fiery A. Cushman, Lisa E. Stewart, Kelly Lowenberg, Leigh E. Nystrom, and Jonathan D. Cohen. 2009. "Pushing Moral Buttons: The Interaction Between Personal Force and Intention in Moral Judgment." *Cognition* 111 (3): 364–71.
- Greene, Joshua, and Jonathan Haidt. 2002. "How Does Moral Judgment Work?" *Trends in Cognitive Sciences* 6 (12): 517–23.
- Haidt, Jonathan. 2001. "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment." *Psychological Review* 108 (4): 814–34.
- Hamilton, William D. 1964. "The Genetical Evolution of Social Behaviour." *Journal of Theoretical Biology* 7 (1): 1–52.
- Harman, Gilbert. 1977. *The Nature of Morality*. Oxford: Oxford University Press.
- Hegel, Georg W. F. 1991. *Elements of the Philosophy of Right*. Edited by Allen W. Wood. Cambridge: Cambridge University Press.
- Herman, Barbara. 1993. *The Practice of Moral Judgment*. Cambridge, MA: Harvard University Press.
- Hills, Alison. 2008. "Kantian Value Realism." *Ratio* 21 (2): 182–200.
- Horgan, Terence, and Mark Timmons. 1991. "New Wave Moral Realism Meets Moral Twin Earth." *Journal of Philosophical Research* 16: 447–65.
- Huemer, Michael. 2001. *Skepticism and the Veil of Perception*. Lanham: Rowman & Littlefield.
- . 2005. *Ethical Intuitionism*. London: Palgrave Macmillan.
- . 2008. "Revisionary Intuitionism." *Social Philosophy and Policy* 25 (1): 368–92.
- . 2013. "Transitivity, Comparative Value, and the Methods of Ethics." *Ethics* 123 (2): 318–45.
- Jackson, Frank. 1998. *From Metaphysics to Ethics*. Oxford: Oxford University Press.

- Jasienska, Grazyna, Susan F. Lipson, Peter T. Ellison, Inger Thune, and Anna Ziomkiewicz. 2006. "Symmetrical Women Have Higher Potential Fertility." *Evolution and Human Behavior* 27 (5): 390–400.
- Jones, Arnold H. M. 1948. *Constantine and the Conversion of Europe*. London: Hazell, Watson and Viney.
- Joyce, Richard. 2001. *The Myth of Morality*. Cambridge: Cambridge University Press.
- . 2006. *The Evolution of Morality*. Cambridge, MA: MIT Press.
- Kahane, Guy. 2011. "Evolutionary Debunking Arguments." *Noûs* 45 (1): 103–25.
- Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.
- Kain, Patrick. 2004. "Self-Legislation in Kant's Moral Philosophy." *Archiv Für Geschichte Der Philosophie* 86 (3): 257–306.
- . 2017. "Dignity and the Paradox of Method." In *Realism and Antirealism in Kant's Moral Philosophy: New Essays*, edited by Elke Elisabeth Schmidt and Robinson dos Santos, 67–90. Berlin: De Gruyter.
- Kant, Immanuel. 1991. *The Metaphysics of Morals*. [MS]. Edited by Mary Gregor. Cambridge: Cambridge University Press.
- . 1997. *Lectures on Ethics*. [VE]. Edited by Peter Heath and J. B. Schneewind. Cambridge: Cambridge University Press.
- . 1998. *Groundwork of the Metaphysics of Morals*. [GMS]. Edited and translated by Mary Gregor. Cambridge: Cambridge University Press.
- . 1999. *Critique of Pure Reason*. [KrV]. Edited by Paul Guyer and Allen W. Wood. Cambridge: Cambridge University Press.
- . 2002. *Theoretical Philosophy after 1781*. [ÜE]. Edited by Henry E. Allison and Peter Heath. Cambridge: Cambridge University Press.
- . 2015. *Critique of Practical Reason*. [KpV]. Edited by Paul Guyer and Allen W. Wood. Cambridge: Cambridge University Press.
- . 2016. *Lectures and Drafts on Political Philosophy*. [NF]. Edited by Frederick Rauscher. Cambridge: Cambridge University Press.
- Katsafanas, Paul. 2013. *Agency and the Foundations of Ethics: Nietzschean Constitutivism*. Oxford: Oxford University Press.

- . 2018. “Constitutivism about Practical Reasons.” In *The Oxford Handbook of Reasons and Normativity*, edited by Daniel Star, 367–94. Oxford: Oxford University Press.
- Kleingeld, Pauline. 2017. “Contradiction and Kant’s Formula of Universal Law.” *Kant-Studien* 108 (1): 89–115.
- Korsgaard, Christine M. 1996. *The Sources of Normativity*. Cambridge: Cambridge University Press.
- . 2003. “Realism and Constructivism in Twentieth-Century Moral Philosophy.” *Journal of Philosophical Research* 28 (Supplement): 99–122.
- . 2009. *Self-Constitution*. Oxford: Oxford University Press.
- Langton, Rae. 2007. “Objective and Unconditioned Value.” *Philosophical Review* 116 (2): 157–85.
- Lazari-Radek, Katarzyna de, and Peter Singer. 2014. *The Point of View of the Universe*. Oxford: Oxford University Press.
- Little, Margaret. 1994. “Moral Realism II: Non-Naturalism.” *Philosophical Books* 35 (4): 225–33.
- Mabrito, Robert. 2005. “Does Shafer-Landau Have a Problem with Supervenience?” *Philosophical Studies* 126 (2): 297–311.
- Mackie, John L. 1977. *Ethics: Inventing Right and Wrong*. London: Penguin Books.
- McPherson, Tristram. 2012. “Ethical Non-Naturalism and the Metaphysics of Supervenience.” In *Oxford Studies in Metaethics 7*, edited by Russ Shafer-Landau, 205–34. Oxford: Oxford University Press.
- Mele, Alfred. 1993. “History and Personal Autonomy.” *Canadian Journal of Philosophy* 23 (2): 271–80.
- . 1995. *Autonomous Agents: From Self-Control to Autonomy*. New York: Oxford University Press.
- Mikhail, John. 2007. “Universal Moral Grammar: Theory, Evidence, and the Future.” *Trends in Cognitive Sciences* 11 (4): 143–52.
- Miller, Christian. 2009. “The Conditions of Moral Realism.” *Journal of Philosophical Research* 34 (1): 123–55.

- Moore, George E. 1903. *Principia Ethica*. Cambridge: Cambridge University Press.
- Nagel, Thomas. 1979. "Ethics without Biology." In *Mortal Questions*, 142–46. Cambridge: Cambridge University Press.
- Nichols, Shaun. 2004. *Sentimental Rules: On the Natural Foundations of Moral Judgment*. Oxford: Oxford University Press.
- O'Neill, Onora. 1989. *Constructions of Reason: Explorations of Kant's Practical Philosophy*. Cambridge: Cambridge University Press.
- . 2003. "Constructivism in Rawls and Kant." In *The Cambridge Companion to Rawls*, edited by Samuel Freeman, 347–67. Cambridge: Cambridge University Press.
- Paaby, Annalise B., and Matthew V. Rockman. 2013. "The Many Faces of Pleiotropy." *Trends Genet* 29 (2): 66–73.
- Paakkunainen, Hille. 2018. "The 'Just Too Different' Objection to Normative Naturalism." *Philosophy Compass* 13 (2): e12473.
- Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Clarendon Press.
- . 2011a. *On What Matters*. Volumes 1 & 2. Oxford: Oxford University Press.
- . 2011b. *On What Matters*. Volume 3. Oxford: Oxford University Press.
- Persson, Ingmar. 2004. "The Root of the Repugnant Conclusion and Its Rebuttal." In *The Repugnant Conclusion: Essays on Population Ethics*, edited by Jesper Ryberg and Torbjörn Tännsjö, 187–99. Dordrecht: Kluwer Academic Publishers.
- Pigden, Charles. 1993. "Naturalism." In *A Companion to Ethics*, edited by Peter Singer, 421–31. Oxford: Blackwell.
- Plato. 1997. *Plato: Complete Works*. Edited by John M. Cooper. Indianapolis: Hackett Publishing Company.
- Prinz, Jesse J. 2007. *The Emotional Construction of Morals*. Oxford: Oxford University Press.
- . 2008. "Is Morality Innate?" In *Moral Psychology, Volume 1: The Evolution of Morality: Adaptations and Innateness*, edited by Walter Sinnott-Armstrong, 367–406. Cambridge, MA: MIT Press.
- Putnam, Hilary. 1975. "The Meaning of 'Meaning'." *Minnesota Studies in the Philosophy of Science* 7 (1): 131–93.

- Rachels, James. 2019. *The Elements of Moral Philosophy*. 9th ed. Dubuque: McGraw-Hill Education.
- Railton, Peter. 1986. "Moral Realism." *The Philosophical Review* 95 (2): 163–207.
- . 2014. "The Affective Dog and Its Rational Tale: Intuition and Attunement." *Ethics* 124 (4): 813–59.
- Rauscher, Frederick. 2002. "Kant's Moral Anti-Realism." *Journal of the History of Philosophy* 40 (4): 477–99.
- . 2006. "Razão Prática Pura Como Uma Faculdade Natural." *Ethic@* 5 (2): 173–92.
- . 2015. *Naturalism and Realism in Kant's Ethics*. Cambridge: Cambridge University Press.
- Rawls, John. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.
- . 1980. "Kantian Constructivism in Moral Theory." *The Journal of Philosophy* 77 (9): 515–72.
- . 1999. *Collected Papers*. Edited by Samuel Freeman. Cambridge, MA: Harvard University Press.
- . 2000. *Lectures on the History of Moral Philosophy*. Edited by Christine M. Korsgaard and Barbara Herman. Cambridge, MA: Harvard University Press.
- Raz, Joseph. 2003. "Numbers, with and without Contractualism." *Ratio* 16 (4): 346–67.
- Ridge, Michael. 2007. "Anti-Reductionism and Supervenience." *Journal of Moral Philosophy* 4 (18): 330–48.
- Roojen, Mark van. 2015. *Metaethics: A Contemporary Introduction*. New York: Routledge.
- Rosen, Gideon. 2020. "What Is Normative Necessity?" In *Metaphysics, Meaning and Modality: Themes from Kit Fine*, edited by Mircea Dimitru, 205–33. Oxford: Oxford University Press.
- Ross, William D. 1930. *The Right and the Good*. New York: Oxford University Press.
- Rousseau, Jean-Jacques. 1968. *The Social Contract*. London: Penguin Books.

- Ruse, Michael. 2010. "The Biological Sciences Can Act as a Ground for Ethics." In *Contemporary Debates in Philosophy of Biology*, edited by Francisco J. Ayala and Robert Arp, 297–315. Oxford: Wiley-Blackwell.
- Sayre-McCord, Geoffrey. 1986. "The Many Moral Realisms." *Southern Journal of Philosophy* 24 (Supplement): 1–22.
- Scanlon, Thomas M. 1998. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
- . 2014. *Being Realistic About Reasons*. Oxford: Oxford University Press.
- Schönecker, Dieter, and Elke Elisabeth Schmidt. 2017. "Kant's Moral Realism Regarding Dignity and Value. Some Comments on the Tugendlehre." In *Realism and Antirealism in Kant's Moral Philosophy: New Essays*, edited by Elke Elisabeth Schmidt and Robinson dos Santos, 119–52. De Gruyter.
- Schroeder, Mark. 2005. "Realism and Reduction: The Quest for Robustness." *Philosophers' Imprint* 5 (1): 1–18.
- Sensen, Oliver. 2009. "Kant's Conception of Inner Value." *European Journal of Philosophy* 19 (2): 262–80.
- . 2011. *Kant on Human Dignity*. Berlin: De Gruyter.
- . 2013. "Kant's Constructivism." In *Constructivism in Ethics*, 63–81. Cambridge: Cambridge University Press.
- . 2014. "Universalizing as a Moral Demand." *Estudos Kantianos* 2 (1): 169–84.
- . 2017. "Kant's Constitutivism." In *Realism and Antirealism in Kant's Moral Philosophy: New Essays*, edited by Elke Elisabeth Schmidt and Robinson dos Santos, 197–222. De Gruyter.
- . 2022. "Universal Law and Poverty Relief." *Ethical Theory and Moral Practice* 25 (1): 1–14.
- Shafer-Landau, Russ. 2003. *Moral Realism: A Defence*. Oxford: Oxford University Press.
- Shemmer, Yonatan. 2012. "Constructing Coherence." In *Constructivism in Practical Philosophy*, edited by Jimmy Lenman and Yonatan Shemmer, 159–79. Oxford: Oxford University Press.
- Singer, Peter. 1972. "Famine, Affluence, and Morality." *Philosophy and Public Affairs* 1 (3): 229–43.

- Slingerland, Edward. 2014. *Trying Not to Try: The Art and Science of Spontaneity*. New York: Crown Publishers.
- Smith, Michael. 1994. *The Moral Problem*. Oxford: Blackwell.
- Stern, Robert. 2012. *Understanding Moral Obligation: Kant, Hegel, Kierkegaard*. Cambridge: Cambridge University Press.
- Steven, Pinker. 2002. *The Blank Slate: The Modern Denial of Human Nature*. New York: Viking.
- Street, Sharon. 2006. "A Darwinian Dilemma for Realist Theories of Value." *Philosophical Studies* 127 (1): 109–66.
- . 2008. "Constructivism About Reasons." In *Oxford Studies in Metaethics 3*, edited by Russ Shafer-Landau, 207–45. Oxford: Oxford University Press.
- . 2010. "What Is Constructivism in Ethics and Metaethics?" *Philosophy Compass* 5 (5): 363–84.
- . 2012. "Coming to Terms with Contingency: Humean Constructivism About Practical Reason." In *Constructivism in Practical Philosophy*, edited by James Lenman and Yonatan Shemmer, 40–59. Oxford: Oxford University Press.
- Sturgeon, Nicholas L. 1985. "Moral Explanations." In *Ethical Theory 1: The Question of Objectivity*, edited by James Rachels, 229–55. Oxford: Oxford University Press.
- . 2006. "Ethical Naturalism." In *The Oxford Handbook of Ethical Theory*, edited by David Copp, 91–121. Oxford: Oxford University Press.
- Tannsjö, Torbjorn. 2002. "Why We Ought to Accept the Repugnant Conclusion." *Utilitas* 14 (3): 339–59.
- Temkin, Larry S. 1987. "Intransitivity and the Mere Addition Paradox." *Philosophy and Public Affairs* 16 (2): 138–87.
- . 2012. *Rethinking the Good: Moral Ideals and the Nature of Practical Reasoning*. Oxford: Oxford University Press.
- Timmons, Mark. 2003. "The Limits of Moral Constructivism." *Ratio* 16 (4): 391–423.
- Unger, Peter. 1996. *Living High and Letting Die: Our Illusion of Innocence*. New York: Oxford University Press.

- Vavova, Katia. 2014. "Debunking Evolutionary Debunking." In *Oxford Studies in Metaethics 9*, edited by Russ Shafer-Landau, 76–101. Oxford: Oxford University Press.
- Velleman, David. 2004. "Replies to Discussion on the Possibility of Practical Reason." *Philosophical Studies* 121 (3): 277–98.
- . 2006. *Self to Self: Selected Essays*. Cambridge: Cambridge University Press.
- . 2009. *How We Get Along*. Cambridge: Cambridge University Press.
- Waal, Frans de. 2014. "Natural Normativity: The 'Is' and 'Ought' of Animal Behavior." *Behaviour* 151 (2–3): 185–204.
- Walden, Kenneth. 2018. "Practical Reason Not as Such." *Journal of Ethics and Social Philosophy* 13 (2): 125–53.
- Williams, Bernard. 1981. *Moral Luck*. Cambridge: Cambridge University Press.
- Wilson, Edward O. 1975. *Sociobiology: The New Synthesis*. Cambridge, MA: Belknap Press.
- Wong, David. 2006. *Natural Moralities: A Defense of Pluralistic Relativism*. Oxford: Oxford University Press.
- Wood, Allen. 2008. *Kantian Ethics*. Cambridge: Cambridge University Press.

BIOGRAPHY

Caner Turan is a doctoral candidate in the department of philosophy at Tulane University, working on metaethics and applied ethics. He is the author of “Necessary Constructivism in Kant’s Moral Theory” in the *Philosophy of Kant*, Ricardo Gutiérrez Aguilar Ed. (2019), “Does Autonomous Moral Reasoning Favor Consequentialism?” (*Estudios de Filosofía*, 65, 89–111), and “Are Ambitious Evolutionary Debunking Arguments Self-Refuting?” (*Southwest Philosophical Studies*, 43, forthcoming).