



UMEÅ UNIVERSITY

Mistaken Morality?

An Essay on Moral Error Theory

Emma Beckman

Department of Historical, Philosophical and Religious Studies
Umeå 2018

Umeå Studies in Philosophy 12

© Emma Beckman 2018

Series editor: Pär Sundström
Department of Historical, Philosophical and Religious Studies
Umeå University
SE-901 87 Umeå, Sweden

This work is protected by the Swedish Copyright Legislation (Act 1960:729)
Dissertation for PhD
ISBN: 978-91-7601-847-7
ISSN: 1650-1748
Front cover: Photo by Emma Beckman, 2011.
Electronic version available at: <http://umu.diva-portal.org/>
Printed by: UmU Print Service Umeå University,
Umeå, Sweden 2018
Distributor: Department of Historical, Philosophical and Religious Studies,
Umeå University, SE-901 87 Umeå, Sweden

For Tony

Abstract

This dissertation explores arguments and questions related to *moral error theory* – the idea that morality inevitably involves a fundamental and serious error such that moral judgments and statements never come out true. It is suggested that the truth of error theory remains a non-negligible possibility, and that we for this reason should take a version of moral fictionalism seriously.

I begin by defining error theory as the claim that moral judgments are beliefs with moral propositions as content, moral utterances are assertions of moral propositions, and no positive moral proposition is true. Second, after giving an account of J.L. Mackie’s error theory, I argue that neither Richard Joyce’s nor Jonas Olson’s argument for error theory gives us strong reasons to believe it. According to Joyce, moral discourse presupposes non-institutional desire-transcendent reasons and non-institutional categorical requirements. I challenge this claim by arguing that morality can be understood as an institution, and that the assumption that there are non-institutional moral reasons and requirements can be understood as entering pragmatically into moral conversations. According to Olson, moral discourse involves a commitment to irreducibly normative favoring relations between facts and actions. I challenge this claim by challenging Olson’s response to Stephen Finlay’s argument against absolutist accounts of moral discourse.

Third, I discuss two objections to error theory, and argue that neither gives us strong reasons to reject it. According to the first objection, which is suggested by Terence Cuneo, error theory entails *epistemic error theory*, which has problematic consequences. After indicating some possible responses on part of the epistemic error theorist, I challenge the entailment claim by defending Hilary Kornblith’s account of epistemic reasons as hypothetical reasons. According to the second objection, error theory entails *normative error theory*, which cannot be believed. Although he does not defend this objection, Bart Streumer has given an argument for the unbelievability claim. I challenge Streumer’s argument by suggesting that we might have hypothetical reasons to believe normative error theory and that, properly understood, Streumer’s conclusion is not as radical as it may first appear.

Fourth, I discuss what practical implications the discovery that error theory is true would have for first-order moral thinking and discourse. I argue that if this practice is overall non-morally valuable to us, we ought to revise

engagement in it on the model of role-playing in *live action role-playing games* if we find out that error theory is true. Some have claimed that Richard Joyce's *fictionalism* encounters (*prima facie*) problems. I argue that by incorporating the suggestion that engagement in revised moral practice is modeled on role-playing, fictionalism can escape these problems and preserve the benefits of first-order moral practice.

Keywords: moral error theory, Mackie, Joyce, Olson, normative institution, fictionalism, normative error theory, epistemic error theory, irreducible normativity, abolitionism, conservatism, propagandism.

Acknowledgements

A wise person once said that the greatest gift you can give someone is a portion of your time. During my years as a doctoral student, I have been spoiled.

First and foremost, I would like to express my deepest gratitude to my advisor, Gunnar Björnsson, for all the time and energy he has spent on sharing his philosophical expertise with me. I am grateful for all his insightful and constructive feedback and suggestions, both on my manuscript and on my dissertation project as a whole. I would also like to thank him for believing in this project, and for his meticulousness, patience and perseverance, without which I am certain that this dissertation would never have been finished.

I would also like to express my gratitude to Sten Lindström and Pär Sundström. Sten was my advisor up and until his retirement in 2012 and has since then been my assistant advisor. Pär has been my assistant advisor throughout my time as a doctoral student. I would like to thank them for encouraging me to start writing about error theory, for all their constructive feedback and suggestions, and for keeping in touch after I left Umeå.

During my years in Umeå, I regularly presented drafts of my manuscript at seminars hosted by the department. I would like to thank those who participated with comments and ideas on these occasions. In addition to Gunnar, Sten and Pär, this includes my colleagues Anders Berglund, Petra Green, Kalle Grill, Ebba Gullberg, Peter Karlsson, Lars Lindblom, Jonas Nilsson, Mats Wahlberg, Lars-Daniel Öhman, Jesper Östman, Lars Samuelsson, Bertil Strömberg, Erik Svård-Bäcklin and Inge-Bert Täljedal. Over the years, I have also presented parts of my manuscript at conferences in Gothenburg, Milan, Linköping and Uppsala. I would like to thank everyone who shared their thoughts on my material with me on these occasions.

Ragnar Francén-Olinder and Jonas Olson who have both, at different stages of the writing-process, read my manuscript in its entirety, and discussed it with me at seminars hosted by the department. I am grateful for all their helpful comments and suggestions, and for travelling to Umeå to discuss my manuscript with me. I would also like to thank Jan Andersson, Bo Petersson, Caj Strandberg and Frans Svensson who have, on different occasions, read and commented on parts of my manuscript. Bo was my advisor and teacher at Linköping University. I would like to thank him for encouraging me to continue studying philosophy, and for convincing me that I could do it. I would also like

to thank Jan, who was my teacher at Folkungaskolan in Linköping, for introducing me to philosophy.

In the autumn of 2012 and early spring of 2013, I was a visiting scholar at University of Southern California. I would like to thank Stephen Finlay, Gunnar Björnsson, The Sweden America Foundation, and everyone else who contributed to making this educating and in many ways rewarding experience possible.

Parts of chapters 3 through 6 have been proofread by Semantix. This was made possible by a grant from the department, for which I am grateful.

For making my years in Umeå fun and colorful, I would like to thank Ida Brandt, Karin Bristav, Sandra Dahlin, Jennifer Forssell, Radoslaw Grabowski, Anna Josefson Ekegren and Tomas Karlsson. A special thanks to Tomas for his encouragement during the final months. For staying in touch during my years in Umeå and overseas, I would like to thank Lina Eneberg, Jenny Linde, Hanna Lundin, Marcus Lönebrink, Evelina Marklund and Nils Rutstam. A special thanks to Lina for visiting me in Los Angeles. For all the laughs and good advice, I would like to thank my aunts, my uncle and my cousins. For sharing their wisdom on life, and for their warm hearts, I would like to thank my grandmothers, Greta and Ulla. My mother, Ann-Brit, my father, Ulf, my brother, Carl, and their families, have supported me in each way imaginable during my years as a doctoral student. I am deeply grateful for all their loving encouragement, for checking in on me every day, for all the lunches, and most of all for believing in me when I did not do so myself.

This dissertation is dedicated to Tony. I am grateful for his love and affection, for making me laugh every day, for his antics, for always keeping his cool, and for his constant encouragement and support. But most of all, I am grateful that he has taken me for the greatest adventure of my life.

Emma Beckman
Linköping, April 2018

Contents

Abstract	i
Acknowledgements	iii
1 Introduction	1
1.1 Background	1
1.2 Aim	3
1.3 Outline of the dissertation	4
2 Moral error theory	9
2.1 Error theory.....	9
<i>Cognitivism</i>	10
<i>Failure</i>	14
2.2 Wide error theory.....	25
2.3 Conclusion.....	30
3 The argument from non-institutionality	33
3.1 Mackie's moral error theory	34
3.2 The argument from non-institutionality.....	41
<i>Joyce's argument for the semantic claim</i>	44
<i>Moral practice as an institution</i>	48
<i>Moral authority as a shared background assumption</i>	51
3.3 Conclusion.....	62
4 The argument from irreducibility	63
4.1 The conceptual claim	64
4.2 A challenge to the conceptual claim.....	68
<i>Finlay's relativism</i>	69
<i>Finlay's challenge to absolutism</i>	71
<i>Olson's response to Finlay</i>	73
4.3 Conclusion.....	76
5 The epistemic objection	77
5.1 The entailment claim	78
5.2 Cuneo's challenges to epistemic error theory	80
5.3 A challenge to the entailment claim.....	86
5.4 Conclusion.....	96
6 The psychological objection	99
6.1 The entailment claim	100
6.2 Streumer's argument for the unbelievability claim	101
6.3 Unbelievability reconsidered.....	114

6.4	Conclusion.....	115
7	Against propagandism and conservatism.....	117
7.1	Preliminaries.....	119
7.2	Abolitionism and the value of moral practice.....	120
	<i>The value of moral practice</i>	122
7.3	Against propagandism.....	123
	<i>Three objections to propagandism</i>	124
7.4	Against conservatism	129
	<i>An objection to conservatism</i>	131
7.5	Conclusion.....	137
8	In defense of moral fictionalism	139
8.1	Fictionalism.....	140
8.2	Two prima facie problems for fictionalism	145
8.3	Modeling revised moral practice on LARPs	149
8.4	Handling the prima facie problems	153
8.5	Conclusion.....	163
	References	165
	Sammanfattning.....	175

Chapter 1

Introduction

1.1 Background

In our ordinary everyday lives, we regularly come across moral questions, for example: Are the household chores divided fairly in this family? Can capital punishment be defended from a moral point of view? Do animals have rights? If children have the right to grow up in a safe environment, does this mean that all people have a duty to provide all children with this?

When we think and speak of such issues, we think and speak of things as having moral features. We think and speak of our own and others' *actions and behavior* as morally right or wrong, praiseworthy or blameworthy, obligatory or forbidden, of *people* as morally good or bad, honest or dishonest, virtuous or vicious, of *political and legal policies* as fair or unfair, justified or unjustified, and so on.

The practice of morally recommending and denouncing plays an important role in our lives. It affects our thoughts about what kind of person we want to be or become, our ways of relating to our families, relatives, friends, and others whom we care about, our views on how society is to be governed, and companies are to be run, as well as many other ideas that are important to us. For example: Most of us want to be good parents, and try to teach our children good moral values by telling them to stand up against bullying in school, and attempting to get them to reflect on what it means to be a good friend. Most of us want the household chores to be divided fairly, and feel that we owe it to the environment and future generations to recycle our household waste. At the grocery store, most of us are willing to pay a little more for ethically produced bananas, feel it is our moral responsibility not to buy any unnecessary plastic bags, and hope that the cashier is paid a fair wage at the end of the month.

Our moral beliefs and ideas affect not only our ways of thinking and speaking about ourselves and our own actions and behavior, but also how we think and speak about other people and their actions and behavior, and about political and legal policies. Imagine yourself watching the evening news. Suppose that there is a news story about a woman in Norway who saved six people from a sinking ship by swimming out in freezing water, thereby risking her own life. You would probably think to yourself that what she did was

something extraordinarily praiseworthy. You might even think of her as a hero. Or suppose that there is a report on a Congressman who has accepted bribes from undercover FBI agents. You might think to yourself that it is morally wrong for a man of the state to accept bribes, and conclude that the Congressman is not only unfit for his job, but dishonest or vicious (perhaps both). Alternatively, you might feel sorry for the Congressman, leading you to conclude that entrapment is not a fair policy of catching criminals. Or suppose, finally, that there is an interview with a politician who is motioning to increase tax on sugary foods and drinks (with the purpose of increasing the national health level). You might find this proposal appalling. Perhaps you think that such a tax would restrict people's rights to decide over their own bodies. Alternatively, you might find this proposal sensible. Perhaps you believe that politicians have a duty to help citizens make choices that are in their long-term interests.

Moral error theory (from now on "error theory") is the suggestion that morality inevitably involves a fundamental and serious *error*, such that when we think and speak about things as morally right or wrong, good or bad, obligatory or forbidden, and so on, the thoughts we have and the claims we make inevitably involve a fundamental and serious mistake. More specifically, error theory is the suggestion that moral judgments and claims aim at truth but never succeed in securing the truth.¹

The error theorist is a skeptic. If she is right, the world is not what we take it to be, and some central aspects of how we understand ourselves and the world surrounding us are mistaken. Encountering an error theorist for the first time might render one uneasy, perhaps even slightly perturbed. The error theorist seems to accuse the rest of us of being somehow mistaken or deceived in making moral judgments and claims – indeed in thinking and speaking in moral terms *at all*. This suggestion flies in the face of common sense. Why would anyone think that error theory is true? Is it not obvious that saving people from a sinking ship is a good thing to do, and that it is wrong for politicians to accept bribes? And how can anyone in their right mind deny that those who want to deprive us of our right to decide over our own bodies, are simply wrong?

Error theory is usually motivated by ontological or metaphysical worries concerning the existence or possibility of moral facts and properties. Most error theorists believe that the *function* of moral judgments and claims is to describe aspects of the world. In making moral judgments and claims, we ascribe moral properties to actions, people, and so on. If there are no, or cannot be any, moral facts or properties, any attempt to describe such facts or ascribe such properties will inevitably fail. By offering arguments for the claim that there are no, or cannot be any, moral facts or properties, error theorists attempt to prove that *no moral judgment or claim is true*.

¹ Or at least, that this is the case with some significant subset of these judgments and claims.

Some might find error theory attractive by reflecting on the fact that it seems possible for natural scientists to give a full description of the world without mentioning any moral properties or facts. This error theorist might reason as follows: If the world contained moral properties or facts, we would be able to identify some phenomenon such that it could not be explained without reference to moral properties or facts. But there is no such phenomenon. Moral properties and facts are explanatorily redundant. Consequently, there is no good reason to believe that there are any moral properties or facts.

Others might find error theory attractive after coming to the conclusion that all other metaethical views fail. These error theorists may reason as follows: Either error theory is true, or some version of metaethical realism or metaethical expressivism is true. Neither realism nor expressivism is true. Therefore, moral error theory is true.²

In the contemporary philosophical literature, many scholars are attracted to error theory by reflecting on what is taken for granted in moral thinking and discourse. Many such arguments have focused on the assumptions concerning what types of *reasons for action* moral properties and facts would provide us (moral agents) with.³ This error theorist might reason as follows: If there were morally right or wrong actions, or genuine moral requirements on us to act in some ways rather than in others, it would at least sometimes be the case that someone had moral reasons to act in one way rather than another. But the suggestion that agents have, or can have, such reasons for action cannot be defended. So there are no morally right or wrong actions, or any genuine moral requirements. Different accounts of the nature of moral reasons have been suggested in contemporary philosophical literature, giving rise to different ways of motivating the claim that agents do not have, or cannot have, moral reasons for action.

1.2 Aim

This dissertation investigates one way of developing the idea that morality inevitably involves a fundamental and serious error such that moral judgments and claims never come out true – i.e., one version of error theory. This dissertation is not a defense of error theory. Nor does it attempt to refute it. Instead, it explores some arguments and questions related to the theory – in particular, arguments and questions that concern how it is best formulated, how

² See Joyce (2007:52, 2016a:§4) for these and further examples of how error theory can be motivated.

³ Several scholars, e.g. J. L. Mackie (1977), Richard Garner (1990), Richard Joyce (2001, 2011a) and Jonas Olson (2011a, 2014), suggest that the error of moral thinking and discourse somehow has to do with a special kind of authority, inescapability, or binding force that moral properties and facts are supposed to have. Spelling out this strange – or “queer” – feature in precise terms has proven far from simple. One popular way of attempting to do so, however, is in terms of a strange kind of reasons for action.

it has been motivated, and what implications it has within the field of epistemology and for epistemically relevant states, and for ordinary (first-order) moral thinking and discourse. It is suggested that the truth of error theory remains a non-negligible possibility, and that we for this reason should take (a version of) *fictionalism* seriously.

I start by giving a minimal definition of error theory. On this definition, error theory implies that no positive moral judgment or statement is true. Second, I discuss two well-known arguments for the theory, both of which concern the nature of moral reasons (if there are such reasons). I argue that neither argument gives us strong reasons to believe the error theory. In particular, none of them convincingly shows that moral thought and discourse involve an error of the kind that makes all positive moral judgments and statements untrue. Third, I discuss two objections to error theory, both of which concern what implications (if any) error theory has within the field of epistemology and for epistemologically relevant states and relations. I argue that neither objection gives us strong reasons to believe that error theory is incorrect. In particular, none of them convincingly shows that error theory has implications that the error theorist cannot handle. Fourth, I discuss what implications the discovery that error theory is true would have for ordinary (first-order) moral thinking and discourse. I argue that if such thinking and discourse is overall non-morally valuable to us, we ought to revise engagement in ordinary (first-order) moral practice on the model of role-playing in *live action role-playing games* if we find out that error theory is true. By incorporating this suggestion, fictionalism has potential to preserve the benefits of ordinary moral practice and handle some *prima facie* problems for fictionalism.

1.3 Outline of the dissertation

This dissertation has eight chapters, of which the present is the first. Below is an outline of the topics covered and the arguments defended in the following chapters.

I begin, in chapter 2, by giving the following minimal definition of error theory:

ERROR THEORY Moral judgments are beliefs that have moral propositions as their content, moral utterances are assertions of moral propositions (**COGNITIVISM**), and no positive moral proposition is true (**FAILURE**).

ERROR THEORY implies that no positive moral judgment or statement is true. I provide examples that illustrate that **ERROR THEORY** is entailed by accounts of error theory given by prominent error theorists. I discuss two problems

encountered by some formulations of error theory – viz. that error theory is self-refuting, and that it has (contradictory) first-order moral implications – and explain how ERROR THEORY can handle them.

I contrast ERROR THEORY to theories which do not imply ERROR THEORY but none the less seem to suggest that ordinary moral thinking systematically relies on false assumptions. I suggest that both Friedrich Nietzsche’s moral skepticism and Axel Hägerström’s axiological nihilism implies error theory in this wider sense of the term.

Second, I discuss two well-known arguments for error theory. The first of these is what I refer to as *the argument from non-institutionality*. This argument is the topic of chapter 3. According to this argument, which is suggested by Richard Joyce (2001, 2006, 2011a), moral discourse presupposes non-institutional desire-transcendent reasons and non-institutional categorical imperatives, whereas all *genuine* desire-transcendent reasons and categorical imperatives are institutional (Joyce 2011a:523). I challenge the argument from non-institutionality by formulating two challenges to the first of these claims. First, I challenge the claim that moral discourse presupposes specifically *non-institutional* moral reasons and requirements by exploring the possibility of arguing that, contrary to what Joyce claims, the suggestion that morality is an institution does not go against the way we think of it. Second, I challenge the claim that moral discourse presupposes non-institutional moral reasons and requirements specifically *in a semantic sense* that entails that a positive moral proposition is true only if there are such reasons and requirements. I argue that it is possible that this assumption enters not semantically but *pragmatically* into moral conversations, by figuring as a shared background assumption in moral conversations. If so, it does not affect the truth values of moral statements. Consequently, even granting that all genuine moral reasons and requirements are institutional, it is possible that ERROR THEORY is not true, for it is possible that moral discourse presupposes only *institutional* reasons and requirements. And even granting that moral discourse and thinking typically presupposes non-institutional moral reasons and requirements, it is possible that ERROR THEORY is not true, for it is possible that the presupposition is *pragmatic* in nature.

In chapter 4, I discuss a second argument for error theory, namely what I refer to as *the argument from irreducibility*. According to this argument, which is suggested by Jonas Olson (2014, 2016a, 2016b), moral discourse involves a commitment to *irreducibly normative favoring relations* between facts and actions, while it is far from clear that there are, or can be, such relations (Olson 2016a:400-1). According to Olson, moral claims are or entail claims about irreducibly normative reasons. He defends this (semantic) claim by responding to Stephen Finlay’s (2006, 2008) relativistic account of moral discourse. Olson challenges Finlay’s argument against absolutist (i.e., non-relativist) accounts of moral discourse, and argues that evolutionary considerations favors the

suggestion that moral claims are or entail claims about irreducibly normative reasons over Finlay's suggestion that moral statements are rhetorical devices used to put pressure on people's actions. I challenge the argument from irreducibility by challenging Olson's response to Finlay. Even granting that irreducibly normative favoring relations are or would be queer, moral claims do not entail claims about queer relations unless they entail claims about irreducibly normative favoring relations, and it is not clear that they do.

Third, I discuss two objections to error theory. The first of these is what I refer to as *the epistemic objection*. This objection is the topic of chapter 5. According to this objection, which is suggested by Terence Cuneo (2007), moral error theory entails an epistemic analogue to it – *epistemic error theory*. The entailment claim can be motivated as follows: Moral error theorists argue that (positive) moral statements are true only if moral reasons have property *f*, and that moral reasons do not, or cannot, have *f*. But (positive) *epistemic* statements can be true only if *epistemic* reasons have *f*. Consequently, if the arguments for moral error theory are successful, epistemic error theory is true. Epistemic error theory is claimed to have several problematic consequences, including that there can be no good arguments for error theory, that error theory cannot be rationally believed, and that error theory is polemically toothless in metaepistemological and metaethical debates (Cuneo 2007:117-122). Consequently, if the epistemic objection is successful, moral error theory has problematic consequences. Moreover, if the arguments for moral error theory show that epistemic error theory is true, and epistemic error theory is false, the arguments for moral error theory are no good; they prove too much. I raise two challenges to the epistemic objection. First, I indicate some ways in which error theorists may attempt avoid some of the problematic consequences of epistemic error theory. Second, I challenge the entailment claim by defending Hilary Kornblith's (1993) account of epistemic reasons. According to this account, epistemic reasons are *hypothetical* reasons, and therefore relevantly *different* than moral reasons. If this is correct, the arguments for moral error theory can be successful without establishing the truth also of epistemic error theory.

In chapter 6, I discuss a second objection to error theory, namely what I refer to as *the psychological objection*. According to this objection, moral error theory entails error theory with respect to all normativity – *normative error theory*. Normative error theory is claimed to have the problematic consequence that it cannot be believed. Consequently, if the psychological objection is successful, moral error theory has entailments that cannot be believed. Moreover, it appears that the unbelievability claim entails *error theory with respect to (thoughts and discourse about) belief in normative error theory*: it appears that if this claim is true, all beliefs and claims to the effect that someone believes normative error theory are systematically in error. Bart Streumer (2013, 2016a, 2016b), who does not defend the psychological objection, and does not intend to raise any other objection to error theory, has argued in favor

of the unbelievability claim.⁴ His argument runs as follows: If normative error theory is true, there are no normative properties, and therefore no reasons to believe normative error theory. We cannot fail to believe what we believe to be entailed by one of our own beliefs, and cannot have a belief while believing that there is no reason for this belief. I challenge the psychological objection by challenging this argument. First, I suggest that even if we cannot have any normative reasons to believe normative error theory, we typically have hypothetical reasons to follow epistemic norms, and might therefore have hypothetical reasons to believe normative error theory. Second, I suggest that when “belief” is understood in a way that makes the premises of the argument plausible, its conclusion is not as radical as it may first seem, and does not imply error theory with respect to belief in normative error theory.

Fourth and last I discuss the practical implications of error theory for ordinary (first-order) moral thinking and discourse. This is the topic of chapters 7 and 8. I suggest that if such thinking and discourse is overall non-morally valuable to us, we ought to carry on with morality as a fiction if we find out that error theory is true. According to fictionalism, which has been suggested by Richard Joyce (2001, 2005), carrying on with morality as a fiction involves revising ordinary (first-order) moral practice in such a way that it no longer involves any *genuine* moral beliefs or assertions, but only moral *pretense-beliefs* and *pretense-assertions*. It has been suggested that fictionalism encounters some (*prima facie*) problems. I argue that fictionalism can handle these problems and preserve the benefits of present moral practice by incorporating the suggestion that engagement in moral pretense-practice is modeled on role-playing in *live action role-playing games*. My argument for this claim has two steps. First, in chapter 7, I challenge alternative answers to the question what we ought to do if we find out that error theory is true that are given in the contemporary philosophical literature – viz. what I refer to as *abolitionism*, *propagandism* and *conservationism*. Second, in chapter 8, I argue that role-playing morality on the model of role-playing in *live action role-playing games* can preserve the benefits of present moral practice, and that by incorporating this suggestion, fictionalism has potential to preserve the relevant benefits and handle the *prima facie* problems.

⁴ In defending the claim that normative error theory cannot be believed, Streumer does not intend to raise an objection to normative error theory. On the contrary, he suggests that our inability to believe normative error theory undermines many objections to it, and that moral error theorists may respond to objections according to which moral error theory implies normative error theory by endorsing his arguments (Streumer 2013:195). Still, if moral error theory entails normative error theory, the claim that normative error theory cannot be believed (at least *prima facie*) seems to entail that error theorists who take themselves to believe (normative or moral) error theory suffer from bad faith or worse. I suggest in chapter 6 that this impression is mistaken: properly understood, the unbelievability claim is not as radical as it may first seem.

Chapter 2

Moral error theory

This chapter provides a minimal definition of moral error theory (from now on: “error theory”) as it is usually understood in the contemporary philosophical literature, and makes some distinctions that will be practical for the discussion in the following chapters. It is suggested that the truth of what will be referred to as “cognitivism” and “failure” is necessary and sufficient for the truth of error theory thus understood. Two related problems connected to the definition of error theory are discussed. It is argued that the definition provided can handle these problems. Some examples of accounts of error theory from prominent defenders of the theory are provided. It is suggested that error theory as defined here is entailed by these accounts. Towards the end of the chapter, error theory is further clarified by being contrasted to theories which do not imply error theory, but seem to suggest that ordinary moral thinking relies, in a systematic way, on false assumptions.

2.1 Error theory

This section provides a definition of error theory and its constituent claims, illustrates how accounts of error theory from some prominent error theorists entail error theory thus understood, and argues that error theory thus understood has potential to handle two problems connected to the definition of error theory.

Error theory can be introduced as the claim that (first-order) morality inevitably involves a fundamental and serious mistake – *an error*.¹ There are several imaginable ways of spelling out this basic idea. The idea most common in the philosophical literature is that the error of morality consists in the fact that the type of thinking and discourse we engage in when we engage in *ordinary (first-order) moral practice* – i.e., that practice of making, defending and arguing for moral judgments and claims which we are all familiar with from our ordinary everyday lives – has truth aspirations but systematically fail to

¹ For a general characterization of the distinction between first-order and second-order moral questions, see Miller (2013:1-3).

secure the truth (Joyce 2011a:519).² Given this, being an error theorist is centrally a matter of accepting this analysis of ordinary moral thought and discourse, and error theory implies that such thinking and discourse suffers from systematic untruth.

Error theory is usually characterized as some combination of (moral) cognitivism, (moral) anti-realism, and what I refer to as “failure theory (with respect to moral truth)”. Different views on how these claims should be understood, as well as on what combination of them is necessary and jointly sufficient for error theory have been suggested in the literature. Some take error theory to necessarily involve all three claims. Others seem to identify it merely with failure theory. A third option – to which I am attracted – is to identify error theory with the conjunction of cognitivism and failure theory. More specifically, I define error theory as follows:

ERROR THEORY Moral judgments are beliefs that have moral propositions as their content, moral utterances are assertions of moral propositions (COGNITIVISM), and no positive moral proposition is true (FAILURE).

To fully understand ERROR THEORY and its implications, we take a closer look at its constituent claims, COGNITIVISM and FAILURE. Spelling out these claims is the aim of the following two sub-sections.

Cognitivism

I said above that according to error theory, the type of thinking and discourse we engage in when we engage in ordinary (first-order) moral practice has truth aspirations but systematically fail to secure the truth. COGNITIVISM is a claim about what we are up to when we think such thoughts and engage in such discourse. More specifically, it is a claim about what type of psychological or mental states moral judgments (primarily) are, and what type of speech acts moral utterances (typically) are:³

² I use “ordinary moral practice” (or simply “moral practice”) to refer to that (first-order) practice of making, defending and arguing for moral judgments and claims which we are all familiar with from our everyday lives. This account of what moral practice is is vague, and intentionally so. I do not intend to offer any precise criteria for what counts as moral practice. Engaging in moral practice involves uttering moral sentences and making of moral judgments. We will see below that if *cognitivism* is true, it also involves making moral assertions and holding moral beliefs.

³ In other words, I understand cognitivism as the combination of what Olson refers to as “psychological cognitivism” and “semantic cognitivism.” The *psychological* question of what mental state one is in when one accepts a moral judgment is logically independent from the *semantic* or *pragmatic* questions of what is expressed by a moral utterance, or what speech act is typically performed by the means of uttering an unembedded moral sentence in the declarative form (Olson 2013:2,9).

COGNITIVISM Moral judgments are beliefs that have moral propositions as their content, and moral utterances are assertions of moral propositions.

Cognitivism is usually contrasted to (*moral*) *non-cognitivism*.⁴ As I use the terms “moral judgment” and “moral utterance”, they are neutral between cognitivist and non-cognitivist interpretations. I use the term “moral judgment” to refer to *the mental state of accepting a moral judgment*. Some have used it to refer to the act of judging that something is morally *f*, or to moral utterances. I use the term “moral utterance” to refer to an utterance of a particular kind of moral sentence, namely an *unembedded moral sentence in the declarative form*. I will use “moral sentence” to refer specifically to unembedded moral sentences in the declarative form, such as “it is wrong to torture animals” or “vegetarianism is morally right”. Moral sentences are to be distinguished from sentences in the interrogative or imperative form, like “did he act wrongly?” or “don’t act wrongly!” and from embedded contexts like “I wonder whether he acted wrongly” and “if it is wrong to torture animals, getting your little brother to torture animals is wrong.” I use the phrases “(an agent) *S* utters a moral sentence” and “*S* makes a moral utterance” interchangeably, to refer to *S*’s tokening specifically of a moral sentence.⁵ Given these ways of understanding the terms “moral judgment”, “moral utterance” and “moral sentence”, COGNITIVISM says that to be in the mental state of accepting a moral judgment is (primarily) to hold a belief (i.e. to be in a cognitive state), and that to utter (or otherwise token) a moral sentence is (typically) to assert a moral proposition. (For ease of presentation, the qualifications “primarily”, “or otherwise token” and “typically” will from now on be taken for granted without being mentioned.)

A belief, as understood here, is a propositional attitude. Propositional attitudes – e.g. believing, wishing, desiring, hoping, doubting – are ways for agents (or subjects) to relate to propositions.⁶ If Jessica hopes that there is pizza in the fridge, she has a hope the content (or object) of which is the proposition that *there is pizza in the fridge*. What Jessica hopes is precisely that this proposition is true (or that the world should be such that this proposition is true). To believe something is to take the attitude of *taking to be the case* or *regarding as true* to a proposition. If Mark believes that *Kennedy was killed by aliens*, he takes it to be the case that *Kennedy was killed by aliens*, or regards this proposition as true.

⁴ For a different account of For accounts of non-cognitivism – which is usually traced back to A. J. Ayer (1936) and C. L. Stevenson (1937) – see Ridge (2013) and van Roojen (2016).

⁵ Any sentence – i.e., any grammatically well-formed unit of language – is a type that may be tokened by speakers. I use “utterance” to refer to such a tokening event.

⁶ More specifically, propositional attitudes are mental or psychological states canonically expressible in the form “*S* *A* that *p*”, where *S* picks out a subject or agent, *A* picks out an attitude, and *p* is a sentence expressing a proposition (see Schwitzgebel 2015).

In addition to being the contents of propositional attitudes, I understand propositions as bearers of truth and falsity (i.e., truth and falsity are properties of propositions). Things that are not themselves propositions – e.g., utterances, sentences and beliefs – cannot be true in themselves, but at most indirectly so. I will say that a belief is true *if and only if* the proposition that constitutes its content is true, and that a sentence is true *if and only if* the proposition that constitutes its content is true.

Given this way of understanding “belief” and “proposition”, COGNITIVISM implies that moral judgments (a type of belief) and moral sentences are the kinds of things that can be evaluated for truth and falsity. They can, in other words, be true or false. A moral judgment is true *if and only if* the moral proposition which is its content is true. A moral sentence is true *if and only if* the moral proposition which is its content is true.

According to COGNITIVISM, to utter a moral sentence is to assert a moral proposition. A different way of phrasing this claim is to say that to utter such a sentence is to make a moral assertion. An assertion is *a speech act in which something is claimed to hold*. What is claimed to hold – i.e., what is “asserted” – is precisely that proposition which constitutes the content of the sentence which is uttered, i.e., of the sentence which is used to perform the speech act. Hence, according to COGNITIVISM, to make a moral utterance is to claim that a moral proposition holds.⁷ Here and below, I use “moral assertion”, “moral statement” and “moral claim” as synonyms.

Moreover, assertions express beliefs. To claim that something, *p*, holds is to express a belief that *p*. The belief expressed by an assertion is, in other words, a belief the object of which is that very proposition which is claimed to hold. This proposition – which is claimed to hold, and which is the object of the belief expressed – constitutes the *content* of the assertion. Given this, COGNITIVISM’s claim that moral utterances are assertions of moral propositions implies that moral utterances express moral beliefs, and that the content of a moral assertion is a moral proposition. I will say that a moral assertion is true *if and only if* the moral proposition which is its content is true.

To understand COGNITIVISM’s claim that moral utterances express moral beliefs correctly, we need to distinguish two notions of “expression”: expression as *a causal or explanatory relation* between a mental state of a speaker and an utterance, and expression as *a property of a type of sentence* independent of

⁷ Speech acts are sometimes referred to as *illocutionary acts*. It is a basic assumption within speech act theory that the communicative significance of an utterance of a meaningful sentence is underdetermined by the (propositional) content of the sentence. To understand the utterance, one must understand both its content and its *illocutionary force*. Suppose, for example, that someone utters the sentence “you will be here tomorrow at noon”. This sentence has a content that is fairly easy to grasp. But to understand how this content is to be taken, one needs to understand with what force the utterance was made. Is it a prediction? A threat? A statement of fact? Terms like “statement” (or “assertion”), “question”, “promise”, “warning” capture different (illocutionary) forces. For more on speech acts, see Green (2017) and Dahllöf (1999:165-173). For more on the speech act of assertion, see Pagin (2016).

the mental states of the speaker (Olson 2013:2). Here, I am interested in the latter notion.⁸ The claim that moral utterances express moral beliefs is a claim about what type of speech acts such utterances are used to perform. More specifically, it is the claim that unembedded moral sentences in the indicative form *conventionally express* (that is, are conventionally used to express) moral beliefs. In contrast, sentences in the interrogative form, like “did he act wrongly?” conventionally express questions. For further illustration of this sense of “express”, consider the following sentence:

Kennedy was killed by Aliens.

Suppose that Mark utters this sentence. Given the present understanding of “express”, it is possible that his utterance *expressed the belief* that Kennedy was killed by aliens even if he did not hold the attitude of taking to be the case towards the proposition that *Kennedy was killed by aliens*. In order to determine whether Mark’s utterance expressed a belief – i.e., whether it was an assertion – we do not examine Mark’s belief-system. Instead, we examine the context where this utterance was made. In particular, we examine what conventions were in force in this context. If, for example, this context is such that there is a convention in force which says that if a speaker utters a sentence, is speaking in a serious tone of voice, is not making a funny or unusual facial expression, is not conditioning her utterance with some expression such as “if it is the case that ...”, etc., then the speaker is to be taken to express a belief, and Mark did make his utterance in a serious tone of voice, did not make any unusual or funny facial expression, did not condition it by some expression such as “if it is the case that ...”, etc., then his utterance expressed a belief (Joyce 2005:296-7).

COGNITIVISM makes two substantial claims – the first about what type of psychological or mental states moral judgments are, the second about what type of speech acts moral utterances are. For the discussion in the following chapters, it is important to keep in mind that the way I use the terms “moral judgment” and “moral utterance”, someone who does not accept COGNITIVISM may accept that speakers engaging in moral discourse regularly make moral judgments and moral utterances.

As readers familiar with the cognitivism versus non-cognitivism debate has probably noticed at this point, I use the terms “belief”, “proposition”, “truth” and “assertion” in robust (as opposed to minimal or deflationary), cognitivist (as

⁸ I do not deny that it is natural to think of belief states as causally co-responsible for assertions. When someone makes an assertion, we interpret her as motivated by the fact that she has a belief that she wishes to communicate to us. For this reason, we interpret her assertion as evidence that she holds the relevant belief (even if she does in fact not hold the belief but is rather out to deceive or lie – it may even be argued that it is due to the fact that we interpret the assertion as evidence of belief that deception or lying is possible). For more on these issues, see Pagin (2016) and Dahllöf (1999).

opposed to non-cognitivist) ways. Someone who uses these and other terms central to my definition of cognitivism in minimal or deflationary ways may (but need not) agree that moral judgments are beliefs, that moral utterances are assertions, or that a moral assertion is true *if and only if* its content is true.⁹

Failure

The second constituent component of ERROR THEORY is

FAILURE No positive moral proposition is true.¹⁰

According to error theory, recall, the type of thinking and discourse we engage in when we engage in ordinary moral practice has truth aspirations but systematically fail to secure the truth. COGNITIVISM can be understood as explaining what is meant by the claim that moral thinking and discourse has truth aspirations. FAILURE explains why, given COGNITIVISM, a subset of this thinking and discourse fails to secure the truth. If moral judgments are beliefs with moral propositions as content, and a belief is true *if and only if* its content

⁹ Some contemporary non-cognitivists attempt to salvage *talk of* moral statements as “true”, and of there being moral “propositions” and “facts” without taking on the ontological, semantic, and epistemological commitments of cognitivism. One way of doing so is to suggest minimal or deflationary ways of understanding the notions of “truth”, “proposition” and “fact”. Roughly put, minimalism about truth is the claim that the collected instances of the schema

(1) ‘S’ is true *if and only if* S

together imply everything there is to know about truth. Given (1), there is no deep difference between saying “abortion is wrong” and saying “*it is true that* abortion is wrong”. Anyone who knows how to use the first sentence knows how to use the second (Dreier 2004:25-6). Minimalists often claim that the function of the predicate “... is true” is simply to allow us to quantify over sentences we affirm or commit ourselves to affirming – i.e., to allow us to say things like “half of what is written in the Bible is true” (Ridge 2013:7). A non-cognitivist who uses such a notion of truth can agree with the cognitivist that moral sentences may be evaluated for truth and falsity, and even that some of them are true. Similarly, minimalism about propositions is, roughly put, the claim that the collected instances of the schema

(2) ‘S’ expresses the proposition that S

together imply everything there is to know about propositions. The function of the expression “the proposition that...” is simply to allow us to form noun-phrases out of sentences. Because a sentence like “abortion is wrong” can be embedded into a that-clause, there are moral propositions. By using such a notion of proposition, a non-cognitivist can agree with the cognitivist that there are moral propositions. And if there are true moral propositions, the non-cognitivist might add, there are also moral facts (Dreier 2004:26). Some contemporary non-cognitivists go as far as to claim that moral utterances are “assertions” that “express beliefs”. It is a matter of debate whether these attempts succeed, and how the minimalist or deflationist notions of truth, proposition, fact, expression, assertion, belief, etc., utilized by these non-cognitivists differ from those utilized e.g. by traditional moral realists. For helpful discussion, see Dreier (2004), Olson (2013), Ridge (2013) and van Roojen (2016). For a famous (broadly) non-cognitivist attempt to “earn the right to” realist-seeming features of moral discourse, see Simon Blackburn (1984, 1993). (Blackburn calls his view “quasi-realism”.)

¹⁰ The distinction between failure theory and success theory was introduced by Geoffrey Sayre-McCord. Sayre-McCord (1988:9-10) describes success theory as the view that moral claims have truth-values and *some* of them are true, and failure theory as the view that moral claims have truth-values but *none* of them are true. FAILURE merely says that no *positive* moral proposition is true, and is therefore not identical to failure theory so defined. For discussion, see Pigden (2010:27-29) and Joyce (2011b:153).

is true, FAILURE implies that no moral judgment the content of which is a positive moral proposition is true. If moral utterances are assertions with moral propositions as content, and an assertion is true *if and only if* its content is true, FAILURE implies that no moral assertion the content of which is a positive moral proposition is true.

There are several possible states of affairs where FAILURE is true. It is true if all positive moral propositions are false, if all positive moral propositions are neither true nor false, and if some positive moral propositions are false while the rest are neither true nor false.¹¹

To fully understand FAILURE, we need to know what makes a moral proposition “positive”. This qualification is introduced to handle two related problems encountered by some formulations of error theory – viz. that error theory is self-refuting and has first-order moral implications.

The problem of self-refutation arises for error theory when error theory is understood as the conjunction of COGNITIVISM and the claim that

FAILURE* No moral proposition is true.

Consider some moral sentence (i.e., an unembedded moral sentence in the declarative), e.g.

(a) abortion is wrong.

COGNITIVISM implies that (a) is true *if and only if* its content, viz. the proposition that *abortion is wrong*, is true. FAILURE* implies that this proposition is not true. Given the law of the excluded middle, this implies that

(b) abortion is not wrong,

is true.¹² If (b) is true, then there is a true moral proposition, and FAILURE* is false. Because we started from the assumption that error theory understood as the conjunction of COGNITIVISM and FAILURE* was true, this indicates that error theory thus understood is self-refuting, indeed incoherent.¹³

Because this problem is caused by the simultaneous truth of COGNITIVISM, FAILURE*, and the law of the excluded middle, the error theorist needs to reject at least one of these claims to solve it. In my view, her best option is to reject

¹¹ Consequently, FAILURE should not be understood as identical to the claim that *all* positive moral propositions are false. If the latter is true, no positive moral proposition such that it is *neither true nor false*.

¹² *The law of the excluded middle* says that if a proposition, *p*, is not true, *p* is false. This means that if *p* is not true, the negation of *p* – i.e. *not-p* or *it is not the case that p* – is true.

¹³ As Charles Pigden (2010:27-8) points out, this problem arises for anyone who claims that the content of the statements that belong to some area of discourse, *D*, are propositions, and that all such propositions are false. If the negation of a *D*-proposition is also a *D*-proposition, then it simply *cannot* be the case that *all D*-propositions are false.

FAILURE*. More specifically, her best option is to formulate error theory in such a way that it entails merely that no proposition belonging to *some relevant subset* of the moral propositions is true.¹⁴ One way of identifying this subset – to which I am attracted – is to claim that it consists of the totality of the *positive* moral propositions. Given this, error theory is the conjunction of COGNITIVISM and the claim that no positive moral proposition is true (i.e., FAILURE). But what makes a moral proposition positive? One possibility is to base this account on Walter Sinnott-Armstrong’s account of positive moral *facts*:¹⁵

[...] [a moral fact is] positive when and only when its description entails any claim about what it is morally wrong to do or not to do, what anyone morally ought or ought not to do, what is morally good or bad, and so on (Sinnott-Armstrong 2006:35).

Based on this, it could be suggested that

POSITIVE* A moral proposition *p* is *positive if and only if p* itself entails something about what it is morally wrong to do or not to do, what anyone morally ought or ought not to do, what is morally good or bad, and so on.

Given POSITIVE*, a moral proposition *p* is *positive if and only if p* entails that something *has* some specific moral property – *if and only if p* ascribes some moral predicate (directly or by implication) to something. If *p* entails merely that something is not morally wrong, that it is not the case that some agent morally ought to do something, etc., *p* does not ascribe any moral predicate, and is not a positive moral proposition. Upon first examination, it might seem that if the error theorist claims that no positive moral proposition is true, and understands “positive” as defined in POSITIVE*, she can coherently admit that the proposition that *abortion is not wrong* is true, for this is not a positive moral proposition.

This impression is mistaken, however, for POSITIVE* does not take (moral) permissibility into account. Permissibility seems to be a moral property. Moreover, it seems that “permissible” means “not wrong”, for it seems that whenever an action is not wrong, it is permissible, and that whenever an action is permissible, it is not wrong (Olson 2011a:69). If error theory implies that the proposition that *abortion is not wrong* is true, and this implies that the proposition that *abortion is permissible* is also true, the self-refutation problem

¹⁴ One alternative strategy might be to argue that there are propositions that are neither true nor false, and that these constitute exceptions to *the law of the excluded middle*.

¹⁵ Walter Sinnott-Armstrong (2006:33-6) supplies this account in an attempt to solve the incoherence problem understood as concerning whether moral error theorists are committed to the existence of *moral facts and properties*. His preferred solution is that error theorists deny the existence not of *all* moral facts but merely of *positive* moral facts.

reappears for the error theorist who accepts POSITIVE*. The reason for this is that POSITIVE* implies that the proposition that *abortion is permissible* is a positive moral proposition.

It seems to me that the error theorist can escape the problem of self-refutation by understanding error theory as the conjunction of COGNITIVISM and FAILURE, where “positive” is understood as

POSITIVE A moral proposition *p* is *positive if and only if p* itself has implications that one can be in violation of.

POSITIVE links the notion of a positive moral proposition to action by suggesting that a moral proposition *p* is positive *if and only if p* itself has implications that can be practically violated (i.e., that an agent can violate by doing or not doing something). First, POSITIVE suggests that *p* is positive if it implies that something is morally wrong to do (or not do), or that someone morally ought to do (or not do) something. Second, it suggests that *p* is positive if it implies that something is morally good, bad, or neutral, *if* something’s being good, bad, or neutral is taken to essentially imply that it ought to be pursued, ought to be prevented, or ought to be given no weight in deliberation, respectively. Third, POSITIVE suggests that *p* is positive if it implies that something is morally permissible, *if* something’s being permissible is taken to essentially imply that one ought not to be blamed for doing it.

However, “permissible” is ambiguous. In the sense just mentioned, it implies “ought not be grounds for blame” or perhaps “is in accordance with the correct moral norms”. Given this understanding, the proposition that *abortion is permissible* essentially implies that someone who performs an abortion ought not be blamed for doing so, and/or that if someone aims to perform an abortion, there is a duty on others not to hinder this person from doing so.¹⁶ If so, the proposition that *abortion is permissible* has implications that one is in violation of whenever one blames someone for performing an abortion or hinders someone from doing so. POSITIVE suggests that when “permissible” is used in this sense, the proposition that *abortion is permissible* is a positive moral proposition. But in the other sense, “permissible” merely means “not wrong” or “not in violation of what is required of us”. Given this understanding, the proposition that *abortion is permissible* does not express a requirement or anything that implies the existence of a requirement, and so does not have any implications that one can be in violation of. POSITIVE suggests that when

¹⁶ Ronald Dworkin seems to have this sense of “permissible” in mind when he says that in an argument about abortion, the view that abortion is “permissible but not mandatory [...] is not neutral in the argument they are having. It takes a position – it holds [...] [that both the view that abortion is forbidden and the view that it is mandatory are] wrong – and it has a direct implication for action because it holds that it would be wrong to interfere with an abortion decision, either way, on moral grounds (Dworkin 1996:94-5).”

“permissible” is used in this sense, the proposition that *abortion is permissible* is not a positive moral proposition.

Whether the error theorist can escape the problem of self-refutation depends upon whether error theory entails that the moral proposition that *abortion is permissible* is a positive moral proposition.¹⁷ Given POSITIVE, and that “permissible” is ambiguous in the way indicated above, some but not all propositions ascribing permissibility to some action are positive moral propositions. Consequently, the error theorist can admit that it is not true that *abortion is wrong* without committing herself to truth of a positive moral proposition, and (thereby) without rendering her position self-refuting.

Another problem that is often discussed in connection to the problem of self-refutation is that error theory seems to have (*contradictory*) *first-order moral implications*. This problem is independent of how the error theorist understands the term “positive”, for it is a consequence merely of the following assumptions: “permissible” means “not wrong”, “impermissible” means “wrong” (or at least “not permissible”), permissibility and wrongness are moral properties, and error theory implies (at least) that some moral proposition ascribing wrongness or impermissibility to something is not true. If a proposition that ascribes either wrongness or impermissibility to an action, ϕ , is untrue, a proposition that ascribes the contrary moral property to ϕ is true. (If it is not true that abortion is wrong, abortion is permissible. If it is not true that abortion is impermissible, abortion is permissible or not wrong. And so on.) Consequently, error theory has first-order moral implications. This is problematic for error theorists who take error theory to be a strictly second-order view (Olson 2011a:69, 2014:12). According to J.L. Mackie, for example,

[...] [error theory] is a second order view, a view about the status of moral values and the nature of moral valuing, about where and how they fit into the world. [...] [F]irst and second order views are not merely distinct but completely independent: one could be a second order moral sceptic without being a first order one, or again the other way round (Mackie 1977:16).

Moreover, if “permissible” means “not wrong”, and “impermissible” means “not permissible”, then if error theory implies that no proposition that attributes wrongness or permissibility to an action is true, error theory seems to imply *both* that any action is permissible, and that any action is *impermissible*.

¹⁷ The reason why POSITIVE* does not solve the problem of self-refutation is that it implies that the proposition that *abortion is permissible* is a positive moral proposition. Recall: error theory implies that *abortion is wrong* is not true. This entails that *abortion is not wrong* is true, and consequently that *abortion is permissible* is true. Given POSITIVE*, this entails that there is a true positive moral proposition.

Consequently, error theory seems to land in straightforward contradiction (Olson 2011a:69, 2014:12).¹⁸

It seems to me that the error theorist can escape the problem of (contradictory) first-order implications. If “permissible” is ambiguous in the way indicated above, the error theorist can argue that there is a sense of “permissible” given which the claim that φ is permissible does not express a requirement or anything that implies the existence of a requirement.¹⁹ Moreover, she may argue that any context in which she is ready to admit that some positive moral proposition is true is a context where “permissible” is used in precisely this sense.

I take the following reflections to lend further support to the claim that the relation between φ is permissible and φ ought not be grounds for blame (or anything else that implies the existence of a requirement) is not necessary in nature: In general, if A implies B, it *cannot* be the case that A without it being the case that B. For illustration, whenever an object is *square*, it is necessarily the case that this object is *not round*. Because this relation between x is *square* and x is *not round* is both necessary and obvious to us, we would consider it a misuse or sign of confusion if someone would claim about some object that it is square while making it clear that he does not accept that it is not round. We would suspect that the speaker has misunderstood the meaning of “square”, or is using “square” in a derived or non-literal sense, etc. However: if someone would claim that the proposition that abortion is permissible is true while making it clear that he does not accept (e.g.) that it is morally wrong to blame someone from performing an abortion, it is possible that we *would not* take it as a sign of misuse or confusion, for it is possible that the conversational context in which the utterance is made is one where it is part of the common ground that radical moral skepticism (like error theory) is a live contender (Joyce 2013:2). In other words, it is possible that the conversational context is such that we understand that the speaker denies that there are any true moral propositions that *require anything of us*. If so, we understand that he uses “permissible” in a sense that does not express a requirement or anything that implies the existence of a requirement.²⁰

¹⁸ To see this, recall that error theory implies that φ is *wrong* is not true, which implies that φ is *not wrong*. If φ is *not wrong* implies φ is *permissible*, error theory implies that φ is *permissible*. But according to error theory, φ is *permissible* is not true. And if φ is *permissible* is not true implies that φ is *impermissible* is true, error theory implies that φ is *impermissible*.

¹⁹ Olson (2011a:69-70, 2014:14, 2016a:398) challenges the claim that “not wrong” and “permissible” are related by entailment or implication by suggesting that the connection is rather a matter of *generalized conversational implicature*. According to most systems of moral standards, anything that is not wrong according to this standard is permissible according to it, and normally when we claim that something is not wrong, we speak from within such a system.

²⁰ In relation to this, it could be argued that the connection between “permissible” and “ought not be grounds for blame” is a matter of *pragmatic (speaker) presupposition*. It could be argued, that is, that *normally* when we claim that action φ is permissible, we expect it to be common ground among ourselves and our audience that if some action is permissible, performing it ought not be grounds

To sum up, I understand error theory as the conjunction of COGNITIVISM and

FAILURE No positive moral proposition is true,

where “positive moral proposition” is understood as “moral proposition that itself has implications that one can be in violation of.”

According to COGNITIVISM, recall, a moral judgment is true *if and only if* the proposition that is its object is true, and a moral assertion is true *if and only if* the proposition that is its content is true. This implies that there is some true moral judgment or assertion only if some moral proposition is true. Given COGNITIVISM, FAILURE implies that no moral judgment or moral assertion the content of which is a positive moral proposition is true. In other words, there is some *true* positive moral judgment or assertion only if FAILURE is false.

If error theory is identical to the conjunction of COGNITIVISM and FAILURE, it does not necessarily involve any ontological claim about moral properties or facts. However, many who have defended error theory have done so for ontological reasons in general and by denying the existence or possibility of moral properties and facts (somehow conceived) in particular.²¹ Mackie, for example, opens his seminal defense of error theory as follows:

There are no objective values (Mackie 1977:15).²²

Richard Joyce, Jonas Olson and Charles Pigden suggest that error theorists take the fact that there are no moral properties to *explain why* morality involves an error

[...] the moral error theorist maintains that moral judgments are truth-evaluable assertions [...], but that the world doesn't contain the properties (e.g., moral goodness, evil, moral obligation) needed to render moral judgments true (Joyce 2013:1).

[...] standard versions of error theory hold that moral judgements are assertions that attribute mind-independent (but non-instantiated) moral properties to objects and that, as a

for blame. A normal context, in this case, would be context where we are evaluating ϕ from a moral point of view, etc. (For further discussion of pragmatic presupposition, see chapter 3)

²¹ According to Olson (2011a:62), moral error theorists *typically* hold that there are no moral facts.

²² Mackie says that error theory “[...] is an ontological thesis, not a linguistic or conceptual one. It is not [...] a view about the meanings of moral statements. Again, no doubt, if it is to be at all plausible, it will have to give some account of their meanings [...]. But this [...] will be a development of the theory, not its core (Mackie 1977:18).”

consequence, moral judgements are systematically mistaken (Olson 2014:4).

I think (as a first approximation) that moral judgments, specifically moral judgments concerning the *thin* moral concepts (“good,” “right,” “ought,” “wrong,” etc.), are propositions, that they are (in the current jargon) *truth apt*. And I think that they are all false. For there are no such properties as goodness, badness, wrongness or obligatoriness. You can’t do genuinely good deeds since there is no such property as goodness for your deeds to instantiate: at best they can be good in some watered down or ersatz sense (Pigden 2010:18).

Error theorists who claim that there are no moral properties or facts accept

ANTI-REALISM There are no moral properties.²³

Because ANTI-REALISM is vague in the sense that it is neutral with respect to the nature and metaphysics of moral properties, it can be elaborated in different ways, and accepted on different grounds.

A first issue on which there may be disagreement is whether moral facts are “mind-dependent” – i.e., whether x :s having a moral property f depends constitutively on the psychological attitude or response someone has (or would have) towards x (Huemer 2005:2). Whether scholars agree or disagree on this point seems to be independent on whether they accept or deny ANTI-REALISM or ERROR THEORY.

The important idea ANTI-REALISM is intended to capture is that however one conceives of the nature of moral facts, to accept ANTI-REALISM is to deny that there are such facts. If moral facts are taken to be mind-dependent, to accept ANTI-REALISM is to deny that there are any mind-dependent moral facts. If moral facts are taken to be mind-independent, to accept ANTI-REALISM is to deny that there are any mind-independent moral facts. And so on. But to claim, for example, that while most people believe that moral facts are mind-independent, these facts are actually mind-dependent, is not to accept ANTI-REALISM, for the truth of this claim does not entail that there are no moral facts.

²³ This is not an attempt to provide a generally applicable definition of metaethical anti-realism. Scholars disagree on how “realism” and “anti-realism” should be understood. Some claim that the distinction is semantic in nature, others that it is ontological (or metaphysical), yet others that both positions incorporate both semantic and ontological elements. Moreover, the question of what moral properties or facts are is one of the main points of division among scholars in metaethics, and even those who agree on the characterization of this and other elements contained in the definitions of realism and anti-realism may disagree on the formulations of the necessary and sufficient criteria for either claim (see Joyce 2016a).

Another issue on which those who accept ANTI-REALISM may disagree is whether (i) moral properties do not exist, or (ii) moral properties exist but are not instantiated in the actual world (our world). It seems that those who take the truth of ANTI-REALISM to explain why ERROR THEORY is true are best advised to understand ANTI-REALISM along the lines of (ii). The reason for this is that on a popular account of the metaphysics of (structured) propositions, propositions consist of properties and things. This seems to imply that if no moral properties exist, no moral *propositions* exist. But if this is the case, it is hard to make sense of the claim that the fact that there are no moral properties explains why moral propositions are false. Olson, for example, says that he assumes

[...] a liberal account of properties according to which there is a property *F* if there is in some natural language a meaningful predicate that purports to pick out *F* [...]. The predicate ‘morally wrong’ fits this description, so there is a property of moral wrongness but error theorists maintain that this property is not instantiated (Olson 2014:2).²⁴

Although many error theorists accept ANTI-REALISM, this claim is best understood not as a component of the theory itself, but as a something that might explain why it is true. The reason for this is that the truth of COGNITIVISM and FAILURE is *sufficient* for it to be the case that morality inevitably involves a fundamental and serious error. These claims imply that a significant subset of our moral judgments and assertions – viz. those that are positive – aim at truth but uniformly fail to secure it. If no positive moral proposition is true, the existence of moral facts seems either ruled out or irrelevant. For this reason, there is no need to add ANTI-REALISM to the definition of error theory.

In addition to being sufficient for the truth of error theory, COGNITIVISM and FAILURE are both *necessary* for its truth. To my knowledge, no one has suggested that COGNITIVISM is sufficient for error theory (it needs to be combined with ANTI-REALISM, FAILURE, or something similar). But it is thinkable that someone should suggest that FAILURE, or something similar, is sufficient. Chris Daly and David Liggins (2010:216) understand error theory as a claim about the truth-values of sentences, and suggest that

[t]o be an error theorist of a discourse is to claim that none of its sentences is true (ibid:209).

²⁴ Those who accept the claim that *moral properties exist but are not instantiated in our world* may disagree on whether it is necessarily or contingently the case that moral properties are not instantiated. Hence, the claim that moral properties *exist* is consistent with the claim that they are *necessarily not instantiated in the actual or any possible world* (e.g. because they are too queer to be instantiated) (Olson 2014:2).

Supposing that the set of *moral* sentences includes at least the set of unembedded moral sentences in the declarative form, this passage can be taken to suggest that *moral* error theory is identical to the claim that

- (1) no moral sentence is true.²⁵

Someone who accepts (2) probably also accepts the following claim (where “moral judgment” refers to the mental state of accepting a moral judgment)

- (2) no moral judgment is true.

Neither of the claims (1) and (2) is implied by FAILURE, for FAILURE (when combined with COGNITIVISM) implies merely that no *positive* moral judgment or sentence is true. It seems to me that (1), (2), and FAILURE have in common that their truth is not *in itself* sufficient for the truth of error theory. First, either claim can be true without it being the case that moral judgments or utterances *aim at truth*. But if moral judgments and utterances do not aim at truth, the fact that they fail to secure it can hardly be referred to as an “error”. Second, if (1), (2), or FAILURE is sufficient for error theory, anyone who rejects COGNITIVISM (and does not suggest some minimalist or deflationary conception of truth, proposition, etc.) is an error theorist. If moral thinking and discourse contains no element that can become subject to evaluation in terms of truth and falsity, it is *a priori* true that no moral judgment or use of a moral sentence in moral discourse is true.²⁶ This is a counterintuitive result.

Moreover, if being in the mental state of accepting a moral judgment does not involve the attitude of *taking to be the case* or *regarding as true* with respect to some proposition, it makes no sense to claim that because such judgments are uniformly untrue, we are making a fundamental and serious mistake when we engage in moral thinking. Similarly, if using moral sentences in moral discourse does not involve *asserting propositions* or *expressing beliefs* (or anything else that can be evaluated for truth or falsity), it makes no sense to claim that because such utterances are uniformly not true, we are making a fundamental and serious mistake when we engage in moral discourse. It is not clear what either of these mistakes would consist in. On the contrary, the opposite is true: if moral utterances do not conventionally express moral beliefs, attempting to use a moral sentence to express a moral belief is an attempt at misuse.

²⁵ My aim of discussing these examples is pedagogy, not exegesis. I am not suggesting that this is a correct representation of how Daly and Liggins define error theory here or elsewhere.

²⁶ Daly and Liggins (2010:211) claim that their definition of error theory *does not exclude* non-cognitivist expressivists from being error theorists. I am suggesting that if understood as (1), their definition has a stronger implication.

If this is correct, neither (1), (2) nor FAILURE is in itself sufficient for error theory, for it is possible that no moral judgment or utterance is true without it being the case that moral discourse and thinking inevitably involves a fundamental and serious mistake.

I take ERROR THEORY to constitute a minimal definition of error theory as usually understood in the philosophical literature. The following passages from Richard Joyce, Jonas Olson and Richard Garner – all of whom have defended error theory – seem to imply the truth of ERROR THEORY:²⁷

An error theory [...] may be characterized as the position that holds that a discourse typically is used in an assertoric manner, but those assertions by and large fail to state truths (Joyce 2001:9).

The moral error theorist thinks that moral language and moral thinking aim at the truth (i.e., that moral language is assertoric and moral thinking doxastic) but that they systematically fail to secure it (Joyce 2011a:519).

[...] moral error theory [...] [is] the view that moral thought and discourse involve systematically false beliefs and that, as a consequence, all moral judgments, or some significant subset thereof, are false (Olson 2017:58).

In contrast to noncognitivists, Mackie claimed that moral judgments have a truth-value, and in contrast to moral realists and ethical naturalists, he held that each and every one of them is false (Garner 1990:137).²⁸

Understanding error theory as ERROR THEORY allows us to discuss error theory without restricting the focus to any *specific* error theory suggested in the literature. This is practical when discussing what implications (if any) error theory has within the field of epistemology and for epistemologically relevant states and relations (chapters 5 and 6), and what we ought to do if we find out that error theory is true (chapters 7 and 8). The following chapters also discuss specific error theories – in particular, those of Mackie, Joyce and Olson (chapters 3 and 4). I understand these error theories as versions of ERROR THEORY – i.e., as implying ERROR THEORY.

While ERROR THEORY is silent with respect to why COGNITIVISM and FAILURE are true, Mackie, Joyce and Olson provide arguments for error theory. We have already seen that Mackie argues that *there are no objective values*.

²⁷ See also the above quotes from Joyce (2013:1), Olson (2014:4) and Pigden (2010:18).

²⁸ For an account of how Mackie's claims imply ERROR THEORY, see chapter 3.

Joyce argues that moral discourse presupposes *non-institutional authority*, and Olson that it presupposes *irreducible normativity*. These arguments for error theory can be understood as arguments for ERROR THEORY, for if either of them is successful, ERROR THEORY is true.

2.2 Wide error theory

This section further clarifies ERROR THEORY by contrasting it to other theories similar to it. In particular, ERROR THEORY is contrasted to two theories which seem to suggest that morality rests upon a mistake or error but which do not seem to imply that ordinary moral thought and discourse has truth aspirations but systematically fail to secure the truth – viz. Friedrich Nietzsche’s moral skepticism and Axel Hägerström’s axiological nihilism.²⁹ It is suggested that both views imply error theory in this wide, less strict sense of the term.³⁰

I suggest that to accept error theory in a wide, less strict sense of the term is to accept

WIDE ERROR THEORY Ordinary moral thinking systematically relies on false assumptions.³¹

Intuitively, if ordinary moral thinking relies on false assumptions in a systematic way, morality rests upon a mistake or error. However, it intuitively seems possible for moral thinking to systematically rely on false assumptions *without* it being the case that these assumptions are part of the content of moral judgments, or that these assumptions are presupposed in moral thinking in such a way that their falsity makes moral judgments uniformly untrue.

Suppose, for example, that there are no objective values and that we are unaware of this. If so, it seems possible that moral argumentation and motivation systematically relies on the assumption that there are objective

²⁹ Axel Hägerström (1868–1939) was a central figure in Swedish and Scandinavian academia and public intellectual debate during the first parts of the 20th century (Olson 2014:43). In Sweden and Scandinavia, he is known mainly as an early advocate of non-cognitivist emotivism (Pettersson 2011:55).

³⁰ According to Maudemarie Clark and David Dudrick (2007) and Nadeem J. Z. Hussain (2007), Nietzsche accepted ERROR THEORY (at least at some point). Because WIDE ERROR THEORY does not imply that ERROR THEORY is false, I am not challenging this claim. According to Bo Pettersson (2011), Hägerström’s axiological nihilism does not entail ERROR THEORY. Because WIDE ERROR THEORY does not entail ERROR THEORY, I am not challenging this claim.

³¹ I have no detailed account of what it is for moral thinking to systematically rely on false assumptions. On the conventional sense of the phrase, however, “systematically relies on” is harder won than “contains”, or “involves”. Olson (2014:8 ff.) distinguishes between what he refers to as *standard error theory* and *moderate moral error theory*. A theory is of the latter kind if it “attributes the doxastic error of false beliefs to ordinary moral thinking, but is moderate in that it does not take this doxastic error to render false all moral judgments, or some significant subset thereof, and neither give rise to a presupposition failure (ibid:10).” Olson seems to suggest that a theory *attributes* such a doxastic error to ordinary moral thinking if it implies that “ordinary moral thinking involves systematically false beliefs (ibid).”

values, although this assumption is not part of the content of moral judgments. If so, moral thinking would need to be revised, or given a radically new justification if we were to discover the truth. The content of moral judgments and statements could however be left as it was prior to this discovery.

Let us start by considering Nietzsche's moral skepticism. Nietzsche's project is a *revaluation of values* (Nietzsche 2005:156).³² More specifically, it is

[...] a *critique* of moral values, *the value of these values should itself, for once, be examined* [...] (Nietzsche 2006a:7).

In the preface of *Daybreak*, Nietzsche claims that this book

[...] exhibit[s] a contradiction and is not afraid of it: in this book faith in morality is withdrawn [...] Out of morality! (Nietzsche 2006b:4)

As these passages indicate, Nietzsche uses some broadly *moral* considerations in his revaluation. Consider also the following passage:

‘There are so many dawns that have not begun to shine’ – this *Indian* inscription is written on the doorway to the book [i.e. *Daybreak*]. Where does its author *look* for that new morning[?] [...] In a *revaluation of all values*, in an escape from all moral values, in an affirmation and trust in everything that had been forbidden, despised, cursed until now (Nietzsche 2005:121).

To affirm and trust in everything that has been forbidden, despised and cursed is inconsistent with escaping *all* values, for it seems to amount not merely to *denying* some first-order moral claims, but also to *affirming* some other such claims.³³

There are different views on what distinguishes the morality that Nietzsche wants to escape from that which he affirms.³⁴ Brian Leiter suggests that what (ultimately) distinguishes them is that the former incorporates norms that harm “higher men” by obstructing the development of human excellence. According to him, Nietzsche takes the flourishing of “higher men” – human excellence or greatness – to be intrinsically valuable (Leiter 1995:123, 126-127). The following passage speaks in favor of Leiter's interpretation (when “morality” refers to the morality Nietzsche rejects):

³² See Leiter (1995:114); Foot (2002:88); Hussain (2007:157-8); Pigden (2010:20); Olson (2014:16).

³³ See Schacht (1983:347-9, 466 ff.); Leiter (1995:113-114); (1997:263 ff.); Geuss (1997:3).

³⁴ Defining the precise scope of Nietzsche's critique of morality is one of the standing problems in the interpretation of Nietzsche (Leiter 1995:113).

What if a regressive trait lurked in ‘the good man’, likewise a danger, an enticement, a poison, a narcotic, so that the present *lived at the expense of the future*? Perhaps in more comfort and less danger, but also in a smaller-minded, meaner manner? . . . So that morality itself were to blame if man, as species, never reached his *highest potential power and splendour*? So that morality itself was the danger of dangers? . . . (Nietzsche 2006a:8)³⁵

Richard Schacht, on the other hand, suggests that Nietzsche’s central target is rather morality that is “anti-nature” – i.e., harmful to the preservation and enhancement of *life*. According to him, *life* is the “sole locus of value” for Nietzsche (Schacht 1983:354-359). The following passage speaks in favor of Schacht’s interpretation:

I *condemn* Christianity [...]. The Christian church has not left anything untouched by its corruption, it has made an un-value of every value, [...]. Parasitism as the church’s *only* practice; drinking all the blood, all the love, all the hope out of life with its ideals of anemia and ‘sanctity’; the beyond as the will to negate every reality; the cross as the mark of the most subterranean conspiracy that ever existed, – against health, beauty, against anything well constituted, against courage, spirit, *goodness* of the soul, *against life itself* (Nietzsche 2005:65-6).³⁶

To determine what interpretation best captures what Nietzsche had in mind would be to go beyond the scope of this chapter. Fortunately, doing so is not necessary for our present purposes. On either interpretation, Nietzsche rejects some set of first-order moral claims in favor of others. And regardless of what interpretation we settle for, it seems clear that Nietzsche takes the morality towards which he objects to be if not universally accepted then at least widespread.³⁷ In general, to claim that something needs to be revaluated seems to be to presuppose that the thing in question is or has been imperfectly valued. Consequently, it seems to me that on either interpretation, Nietzsche’s project of revaluation presupposes that ordinary moral thinking relies, in a systematic way, on false assumptions about *what is genuinely valuable*, or at least about *the genuine values* of the things that are valuable. Having discovered that these assumptions are false, Nietzsche finds it necessary to radically revise

³⁵ This passage (although differently translated) is quoted by Leiter (1995:127).

³⁶ This passage (although differently translated) is partially quoted by Schacht (1983:357).

³⁷ According to Schacht, Nietzsche considered “the thorough deflation of the pretensions of previously established and commonly accepted moralities and values (Schacht 1983:342)” one of his tasks. According to Leiter, Nietzsche “aims to offer a revaluation of existing values (Leiter 1995:114).”

moral thinking. In other words, it seems to me that Nietzsche's moral skepticism presupposes WIDE ERROR THEORY.

Let us now consider Hägerström's axiological nihilism. In his 1911 inaugural lecture at Uppsala University, Hägerström claims that

The final result of this investigation is that a moral proposition as such – i.e. a proposition to the effect that a certain action represents a supreme value – cannot be said to be either true or false. It is not at all a proposition to the effect that the action is actually or in truth the right one (Hägerström 1964:92).

To understand this claim, we need to consider Hägerström's theory about *psychological acts* (or psychological states, or states of consciousness). Hägerström distinguishes between different kinds of psychological acts in virtue of their different kinds of content. A main division in his theory is between ideas and feelings (from now on, "hägerström-ideas" and "hägerström-feelings"). Hägerström-feelings have contents that can exist only as the content of a psychological act – i.e., that depend for their existence on being experienced by a mind. Hägerström-ideas, in contrast, have contents that can exist in space and time external to the act itself. A third kind of psychological acts are *simultaneous associations*. These are complex psychological acts where (at least) two acts are connected, and their contents are merged (Broad 1951:100; Petersson 1973:30-8, 54-8, 187-8, 2011:56; Olson 2014:45-7).

Two claims Hägerström makes in relation to his act-psychology are particularly important for understanding his claim that a moral proposition as such cannot be said to be either true or false. The first is that *valuations*, including *moral valuations*, are simultaneous associations.³⁸ Suppose that Jane utters the following sentence, which expresses a valuation:

(c) It would be good to have a barrel of potatoes.³⁹

When Jane utters (c), she has a hägerström-idea of being in possession of a barrel of potatoes and a hägerström-feeling of approval towards such a possession. The two acts are connected, and their contents are merged (Broad 1951:100; Petersson 2011:56-7). Second, Hägerström claims that the only psychological acts that can be evaluated for truth and falsity are a subset of the hägerström-ideas. When combined with the claim that moral valuations are simultaneous associations, this implies that moral valuations are psychological acts of a kind that cannot be evaluated for truth and falsity. Consequently, moral

³⁸ Hägerström does not give any precise definition of "valuation". However, it is clear that he (i) distinguishes moral valuations from claims *about* morality, (ii) uses ordinary value predicates ("good", "bad", "right", etc) in his examples of valuations (Petersson 2011:58).

³⁹ This is Hägerström's own example of a sentence expressing a valuation (Petersson 2011:57).

valuations are universally *neither true nor false*, i.e., cannot be said to be either true or false (Hedenius 1940:40; Petersson 1973:188, 2011:56-59; Olson 2014:53).

Another claim Hägerström makes is that we are sometimes mistaken with respect to the contents of our own psychological acts (Petersson 1973:34-6). Taken together with the claim that the content of an act determines what kind the act belongs to, this implies that we are sometimes mistaken with respect to what psychological states we are in. One type of context in which we are so mistaken is when we make valuations. Although the valuation is a simultaneous association, it appears to us that it is a hägerström-idea (Petersson 1973:189, 2011:56-7). Because introspection seems to be the primary (and perhaps also most reliable) way of gaining knowledge about the content of mental states, claiming that introspection is misleading with respect to the nature of valuations comes close to claiming that

- (3) ordinary people have false beliefs about what they are up to when they make moral valuations.⁴⁰

I said above (section 2.1), that *moral* error theory is often understood as the claim that ordinary moral thinking and discourse has truth aspirations but systematically fails to secure the truth. The fact that Hägerström seems to accept (3) might be taken to suggest that he accepts error theory *with respect to thoughts about what we are up to when we make moral valuations*. Suppose that ordinary people do hold beliefs about what they are up to when they make moral valuations. Given this, it seems reasonable to suppose that these beliefs have truth aspirations. But according to (3), these beliefs are false. So it seems that if (3) is true, ordinary people's thoughts about what they are up to when they make valuations aim at truth but systematically fail to secure it.

This being said, it should be noted that (3) can be true without it being the case that ordinary *moral* thinking (including the activity of making valuations) systematically relies on our beliefs about what we are up to when we engage in such thinking, or when we make moral valuations. For analogy, observing that many (most?) people falsely believe that morality presupposes the existence of a god is not in itself sufficient for concluding that their moral thinking systematically relies on false beliefs, for it is possible that this belief does not influence their moral thinking. Given this, the fact that Hägerström accepts (3) is not in itself sufficient for concluding that axiological nihilism implies WIDE ERROR THEORY. It does so only when (3) is supplemented with some claim to the effect that our beliefs about what we are up to when we make

⁴⁰ This is also pointed out by Petersson (1973:189) and Olson (2014:54).

moral valuations influence our moral thinking. It is unclear to me whether Hägerström accepts any such claim.⁴¹

Perhaps, however, there is another ground for concluding that axiological nihilism implies WIDE ERROR THEORY. Bo Petersson points out that Hägerström claims that our moral valuations are accompanied by various kinds of false beliefs, e.g. the belief that moral statements can be proved and that there is, or must be, a god that requires things from us (Petersson 2011:57-8). That moral thinking is *accompanied* by false beliefs does not entail that it *systematically relies on* false beliefs. However, it seems to me that it is possible to interpret Hägerström (1964:92-95) as making also the stronger claim that morality *has been based on* false assumptions. For example, Hägerström says that

Superstitious conceptions have often fallen away – superstitions on which, to a certain extent, the socially and culturally determined popular morality has been based (ibid:94).

He also says that

All the evidence points to the fact that, once we have taken the final step and have left behind any patent or suppressed belief in the cosmic and thus objective significance of value, this popular morality [...] will be born anew [...], with a more emancipated and far-sighted vision. It will also bear the stamp of a milder judgment on the human aspiration, which follows from [...] the insight that [...] nothing is in itself higher or lower (ibid:95).

It seems to me that the fact that Hägerström makes these claims can be taken to suggest that axiological nihilism implies WIDE ERROR THEORY. If morality is based on superstitious assumptions, and it is moreover based on superstitious assumptions *in such a way* that abandoning them will make us start judging differently than we previously did, then it does seem like at least part of our moral thinking systematically relies on these superstitious assumptions. If these assumptions are false, morality relies on false assumptions. If this is correct, Hägerström's axiological nihilism implies WIDE ERROR THEORY.⁴²

2.3 Conclusion

The purpose of this chapter was to provide a definition of moral error theory and to make some distinctions that will be practical for the discussion in the following chapters.

⁴¹ Thanks to Gunnar Björnsson for helpful discussion here.

⁴² Olson (2014:54) draws what is essentially the same conclusion, although for different reasons.

In the first section of the chapter, it was suggested that the truth of COGNITIVISM (moral judgments are beliefs that have moral propositions as their content, and moral utterances are assertions of moral propositions) and FAILURE (no positive moral proposition is true) is necessary and sufficient for the truth of moral error theory as it is usually understood in contemporary philosophical literature. The term ERROR THEORY was introduced to refer to the conjunction of these claims. Some examples of accounts of error theory from prominent defenders of the theory were provided, and it was suggested that these accounts entail ERROR THEORY. Two related problems connected to the formulation of error theory – the problems of self-refutation and (contradictory) first-order implications – were discussed. It was argued that ERROR THEORY can handle these problems.

In the second section of the chapter, ERROR THEORY was contrasted to Friedrich Nietzsche's moral skepticism and Axel Hägerström's axiological nihilism. These theories do not imply ERROR THEORY, but seem to suggest that ordinary moral thinking nevertheless systematically relies on false assumptions.

Chapter 3

The argument from non-institutionality

In this and the following chapter, I discuss two of the most well-known and currently most debated arguments for ERROR THEORY.¹ The first argument, which I refer to as *the argument from non-institutionality*, is suggested by Richard Joyce (2001, 2006, 2011a). The second, which I refer to as *the argument from irreducibility*, is suggested by Jonas Olson (2014, 2016a, 2016b). I argue that while both arguments have their merits, neither gives us strong reasons to believe that ERROR THEORY is true. In particular, none of them convincingly shows that moral thought and discourse involves an error of the kind that makes all positive moral judgments and statements untrue.

The argument from irreducibility is the topic of the next chapter. This chapter discusses the argument from non-institutionality. This argument, which is suggested by Joyce (and possibly by Mackie), says that morality involves an error because moral discourse presupposes *non-institutional desire-transcendent reasons* and *non-institutional categorical imperatives*, while all *genuine* desire-transcendent reasons and categorical imperatives are *institutional* (Joyce 2011a:523). I raise two challenges to this argument. First, I explore the possibility of interpreting moral practice as an institution, differing from other normative frameworks like etiquette and chess not in kind, but only in scope and degree. If so, moral discourse presupposes merely *institutional* requirements and reasons. Second, I suggest that it is possible that the relevant presupposition enters not semantically but *pragmatically* into moral conversations – viz., by figuring as a shared background assumption in moral conversations – so that normally when we make moral statements, we operate under the assumption that all parties to the conversation give practical weight to moral requirements and reasons, and give moral considerations precedence in

¹ We saw in chapter 2 that ERROR THEORY says that moral judgments are beliefs that have moral propositions as their content, moral utterances are assertions of moral propositions (COGNITIVISM), and no positive moral proposition is true (FAILURE). A moral judgment is true if and only if its content is true. A moral assertion (or “statement”) has a moral proposition as its content, and is true if and only if its content is true. ERROR THEORY implies that no positive moral assertion or judgment is true.

practical intra- and interpersonal deliberation. If so, this presupposition does not affect the truth values of moral statements and beliefs.

This chapter begins with a discussion of Mackie's error theory as formulated in his 1977 *Ethics: Inventing Right and Wrong* (from now on: "*Ethics*"). My reason for proceeding in this fashion is threefold. First, Mackie's writings contain traces of both the argument from non-institutionality and the argument from irreducibility. Second, Mackie's error theory highlights distinctions and terminology central to the discussion in this and the following chapters. Third, any exposition of arguments for error theory that does not discuss *Ethics* is to that extent incomplete. It is common knowledge within the canon of Western philosophy that the term "error theory" was coined in this book (Joyce 2016a:§4), which contains some of the historically most discussed arguments for this theory – the arguments from queerness and relativity – and has been described as "one of the landmark texts of 20th-century metaethics (Joyce and Kirchin 2010:x)."

3.1 Mackie's moral error theory

This section discusses Mackie's error theory as formulated in his (1977) *Ethics*.² The aim is not to provide a waterproof reconstruction of Mackie's intended claims and arguments, but to elaborate on some philosophically interesting arguments and distinctions that can be extracted from his writings.³

I proceed as follows: First, I give an account of Mackie's ontological and semantic claims. Second, I give an account of the notion of a categorical imperative, central to Mackie's ontological claim. Third, I explore the possibility of interpreting the arguments from non-institutionality and irreducibility as elaborations of arguments indicated by Mackie. Doing this involves giving some consideration to Mackie's notion of an *objectively valid categorical imperative* and his *argument from queerness*.⁴

² Mackie also defends error theory in his 1946 paper "A Refutation of Morals", and in some passages in his 1980 and 1982 books *Hume's Moral Theory* and *The Miracle of Theism*.

³ The lack of consensus on how Mackie should be understood might be due to the fact that his style of writing is sometimes metaphorical (Dworkin 1996:114) and opaque (Olson 2014:4, 79).

⁴ More specifically, the *metaphysical part* of the argument from queerness is discussed below. Mackie is usually described as giving two arguments for error theory: the argument from relativity and the argument from queerness. The argument from relativity starts from the observation that moral codes and judgments vary with societies, times and even with groups within one and the same society. Mackie argues that this observation does not fit well with the claim that moral judgments are apprehensions of objective truths, but is better explained by the hypothesis that moral norms and judgments reflect variations in ways of life. Mackie is not claiming that the fact that moral views vary in itself proves that there are no objective values, but rather that the variation is more readily explained by the assumption that the context (material, geographical, cultural, etc.) in which a group is situated affects what moral codes are embraced by members of this group (Mackie 1977:36-7, see Olson 2014:72 ff. for discussion). The argument from queerness is usually described as having two parts; one metaphysical and one epistemological (this is how Mackie 1977:38 introduces the argument, see also Joyce 2016a:§4.1), where Mackie offers two kinds of reasons for doubting the existence of objective moral values. The metaphysical part of the argument is an argument for the

Mackie's ontological claim constitutes the opening sentence of the first chapter of *Ethics*:

There are no objective values (Mackie 1977:15).

Mackie expresses the semantic claim (e.g.) as follows:

[...] ordinary moral judgements include a claim to objectivity, an assumption that there are objective values in just the sense in which I am concerned to deny this. And I do not think it is going too far to say that this assumption has been incorporated in the basic, conventional, meanings of moral terms (ibid:35).

It seems plausible to understand Mackie as taking the claim to objectivity he refers to in this passage to be part of the truth-conditions of moral judgments and statements in such a way that the existence of objective values is a necessary condition for the truth of such judgments and statements. Given this, Mackie's claim that the assumption that there are objective values is part of the basic, conventional meanings of moral terms entails that

a moral proposition is true only if there are objective values.

Much discussion of Mackie's error theory (and of error theory in general) has centered on the notion of a "categorical imperative". This is probably due to the following passage, also from *Ethics*:

[...] my thesis that there are no objective values is specifically the denial that any such categorically imperative element is objectively valid (ibid:29).

Given this, it makes most sense to understand the ontological and semantic claims as follows:

(M₁) there are no objectively valid categorical imperatives,

claim that "[i]f there were objective values, then they would be entities of a very strange sort, utterly different from anything else in the universe (Mackie 1977:38)." Objective values would be entities of features "[...] of quite a different order from anything else with which we are acquainted (ibid:40)." (For discussion, see below.) The epistemological part of the argument depends on the metaphysical part. Mackie argues that *because* objective values would be metaphysically peculiar in this way, "if we were aware of them, it would have to be by some special faculty of moral perception or intuition, utterly different from our ordinary ways of knowing everything else (Mackie 1977:38)." Our ordinary ways of acquiring knowledge – sensory perception, introspection, logical inferences, etc. – would be insufficient for acquiring knowledge of objective moral values (ibid:38-9). Olson (2014:84 ff., 2016a:400) suggests that there is a distinction to be made between *the argument from queerness* (singular) and *the queerness arguments* (plural) (see also chapter 4, n. 1 and Shepski 2008).

(M₂) a moral proposition is true only if there are objectively valid categorical imperatives.

Because Mackie is a cognitivist, and his claims (M₁) and (M₂) imply that no moral proposition is true, Mackie accepts ERROR THEORY. But in order to fully understand (M₁) and (M₂), we need to know what *categorical imperatives* are and what it is for them to be *objectively valid*.

Imperatives are requirements. Requirements can be expressed with sentences either in the imperative or in the indicative form – i.e. sentences like (a) or (b), respectively:

- (a) Go to bed now!
- (b) You must to go to bed now.

They can also be expressed with conditional sentences like

If you promised to go to bed before midnight, you have to go to bed now.

Any requirement is either *categorical* or *hypothetical*.⁵ If it is a piece of advice on what the agent should do to satisfy or achieve some desire, end, goal, interest, etc. of hers, it is hypothetical. Otherwise it is categorical (Joyce 2006:60-1).⁶ (When speaking of “desires” in connection to hypothetical imperatives and reasons, I will from now on use “desire” broadly, to cover desires, ends, goals, interests and related states.) Whether an utterance of a sentence expresses a categorical or a hypothetical imperative depends (at least partially) on the force with which it is used (see Joyce and Kirchin 2010:xiv). Consider:

- (c) Don't take the rook.

⁵ This distinction was introduced by Immanuel Kant. According to Kant “[...] all imperatives command either *hypothetically* or *categorically*. The former represent the practical necessity of a possible action as a means to achieving something else that one wills (or that it is at least possible for one to will). The categorical imperative would be that which represented an action as objectively necessary of itself, without reference to another end (Kant 1996:67).” He also claims that “[...] if the action would be good merely as a means to *something else* the imperative is hypothetical; if the action is represented as *in itself* good [...], then it is categorical (ibid.)” Note that while Kant seems to suggest that agents have hypothetical reasons to perform actions that are *necessary as means to their ends* (broadly construed), I am making the weaker claim that agents have hypothetical reasons to perform actions that are *means to their ends* (broadly construed).

⁶ Joyce (2006:61) refers to this analysis of the notion of a categorical imperative as “minimal”. For more on the distinction between hypothetical and categorical imperatives, see Foot (1972), Mackie (1977:28), Joyce (2011a:521-2) and Miller (2013:108).

A chess teacher might use (c) to advise her student that taking the rook would be suboptimal with respect to the desire of winning the game. If so, (c) can be interpreted as (tacitly) conditional – i.e., as expressing the following advice:

If you want to win the game, don't take the rook.

Used in this way, (c) expresses a hypothetical imperative.⁷ It specifies something the student ought to do provided that he has the relevant desire – in this case, to win the game. If it turns out that he does not desire to win the game, the advice no longer applies, and the speaker will (in normal cases) withdraw it. Understood as expressing a hypothetical imperative, (c) is a claim about what the student desires, and what behavior will (or is likely to) promote these desires. As I understood here, hypothetical imperatives are, or can be reduced to, non-normative claims in analogous ways.⁸ I will have more to say about this in chapters 4-6.

Alternatively, (c) might be used to express a requirement of chess (there is a chess rule that forbids taking the rook, and players are required to follow the rules). Chess requirements, etiquette requirements and moral requirements have in common that they are typically expressed with categorical imperatives. Consider:

(d) You must wear black clothes at the funeral.

When used to express a requirement of etiquette, (d) expresses a requirement that applies *even if* none of the addressee's desires would be promoted by wearing black clothes at the funeral, and *even if* some of his desires would in fact be promoted by his *not* wearing black clothes at the funeral. Used in this way, (d) is not offered as advice on how the addressee should act to achieve some of his desires, and the speaker will (in normal cases) *not* withdraw her requirement if she learns that none of the addressee's desires would be served by his wearing black clothes at the funeral. Consequently, when used to express a requirement of etiquette, (d) specifies a categorical requirement – something the addressee ought to do regardless of what his desires are (Mackie 1977:27-8; Joyce 2006:60-2, 2011a:521-2; Joyce and Kirchin 2010:xviii).

Following Joyce (2001:42), I will say that “an agent, *S*, ought to ϕ ” expresses a categorical imperative *if and only if*

if *S* ought to ϕ , then *S* ought to ϕ regardless of whether he cares to, regardless of whether ϕ :ing satisfies any of his desires or furthers his interests.

⁷ Note that hypothetical imperatives need not be explicitly conditioned on agent's desires: it is sufficient that it is *implicitly* so conditioned (see e.g. Joyce 2011a:523).

⁸ For discussion see Olson (2014:152 ff.). For an objection, see Bedke (2010:47 ff.).

Any ought-claim implies a reasons-claim. According to Mackie,

‘Ought’ [...] says that the agent has a reason for doing something (Mackie 1977:66).

[...] a rough general equivalent of ‘a ought to G’ would be ‘There is a reason for a’s G-ing’ (ibid:77).

This idea is supported by the observation that it is intuitively odd for someone to insist that *S* ought to ϕ while denying that *S* has a reason to ϕ (Joyce 2001:38-9).

Some error theorists take categorical imperatives to imply *desire-transcendent* reasons. Mackie claims that:

A categorical imperative [...] would *express a reason for acting* which was unconditional in the sense of *not being contingent upon any present desire of the agent* to whose satisfaction the recommended action would contribute as a means – or more directly: ‘You ought to dance’, if the implied reason is just that you want to dance or like dancing, is still a hypothetical imperative (Mackie 1977:29, my italics).

Joyce similarly claims that any rule-book implicitly contains reasons simply in virtue of being a book of *rules*, and that whenever a rule is expressed as a categorical imperative – e.g. “you cannot take the rook” – it implies a claim expressing a desire-transcendent reason – e.g. “you have a reason not to take the rook regardless of whether doing so satisfies any of your desires, interests, ends, etc.” (Joyce 2001:40, 2011a:522 ff.). Following Joyce (2001:42), I say that *S* has a desire-transcendent reason to ϕ *if and only if*

S has a reason to ϕ regardless of whether he cares to, regardless of whether ϕ :ing satisfies any of his desires or furthers his ends or interests.

Analogously, hypothetical imperatives – e.g. “if you want to win the game, don’t take the rook” – imply *hypothetical* (or “desire-dependent”) *reasons* claims – e.g. “if you want to win the game, you have a reason not to take the rook”. A hypothetical reasons claim is a claim that *S* has a reason to take the means to satisfy his desires, ends, interests, etc. (Joyce 2001:35; Olson 2014:152-4).⁹ I will have more to say about hypothetical reasons in chapters 4-6.

⁹ Further distinctions could be made with respect to the nature of hypothetical reasons. As we saw above (n. 5), some might claim that *S* has a hypothetical reason to ϕ only if ϕ :ing is *necessary* as means to her ends. Joyce (2001:53) distinguishes between what he refers to as “subjective” and “objective” reasons – viz. reasons that the agent is aware of and reasons that she is not aware of. *S*

If categorical imperatives *imply* desire-transcendent reasons claims, a claim “*S* ought to φ ” which expresses a categorical imperative *is true* only if *S* has a desire-transcendent reason to φ – i.e., only if *S* has a reason to φ regardless of whether he cares to, regardless of whether φ :ing satisfies any of his desires or furthers his interests. If the claim that *S* has a desire-transcendent reason (the consequent of the implication) is not true, the categorical imperative claim (the antecedent) cannot be true either.

It is clear that Mackie objects to the existence of objective values and objectively valid categorical imperatives. The challenge is to figure out what he means by “objectivity” as a feature of values, and “objective validity” as a feature of categorical imperatives. To figure this out is to figure out in virtue of what feature of moral values and requirements Mackie takes morality to involve an error, or in other words, what constitutes the target of Mackie’s most acute queerness worry (Olson 2014:117). The arguments from non-institutionality and irreducibility can be understood as representing two competing answers to this question: the argument from non-institutionality suggests that Mackie’s central target is the non-institutionality of moral requirements and reasons, and the argument from irreducibility suggests that it is rather the irreducible normativity of moral reasons.

To see how Mackie’s objection to the existence of objectively valid categorical imperatives contains traces of the argument from non-institutionality, two things should be noted. First, Mackie’s objection is not directed at desire-transcendent reasons and categorical imperatives *per se*, but specifically at *objective* or *objectively valid* such reasons and imperatives. Consequently, Mackie accepts not only hypothetical reasons and hypothetical imperatives, but also any desire-transcendent reasons and categorical imperatives that are *subjective* or *subjectively valid*, whatever subjectivity or subjective validity consists in (Mackie 1977:29; Joyce 2016a:§4.1). Second, while desire-transcendent reasons and categorical requirements are inescapable in the sense that we cannot evade them by referring to our desires, an *objective* desire-transcendent reason or an *objectively valid* categorical requirement would be inescapable *in some further and problematic sense*, whatever sense that is (Garner 1990:138; Joyce 2001:3, 35-7, 2006:60-4, 2011a:522, 2016b:519). Joyce (2001:45, 2011a:523) suggests that this feature of moral reasons and requirements consists in their *non-institutionality*, and makes this feature the

has an *objective* reason to φ if and only if φ :ing will further her ends, and a *subjective* reason to φ if and only if *S* is *justified in believing* that she has an objective reason to φ . Bernard Williams (1981:101) distinguishes between what he refers to as “internal” and “external” reasons – viz. reasons that connect to what Williams (ibid:102) refers to as the agent’s “subjective motivational set” in such a way that it has the potential of motivating her into action and reasons that do not so connect. *S* has an *internal* reason to φ if and only if *S* could, through imaginative reflection and rational deliberation, and armed with true beliefs, come to be motivated to φ . For more on the distinction between internal and external reasons, and the related distinction between (existence forms of) internalism and externalism about normative reasons, see Finlay and Schroeder (2017).

focus of the argument from non-institutionality. According to this argument, morality involves an error because moral discourse presupposes that there can be non-institutional categorical requirements on agents, so that agents can have non-institutional desire-transcendent reasons for action.

To see how Mackie's error theory contains traces of the argument from irreducibility, consider his *argument from queerness*.¹⁰ Mackie claims that

[i]f there were objective values, then they would be entities or qualities or relations of a very strange sort, utterly different from anything else in the universe (Mackie 1977:38).

In the following passage, Mackie elaborates on what the "queerness" of moral values and properties consists in – i.e., on what the world would have to be like in order for it to contain such values and properties:

Plato's Forms give a dramatic picture of what objective values would have to be. The Form of the Good is such that knowledge of it provides the knower with both a direction and an overriding motive; something's being good both tells the person who knows this to pursue it and makes him pursue it. An objective good would be sought by anyone who was acquainted with it, not because of any contingent fact that this person, or every person, is so constituted that he desires this end, but just because the end has to-be-pursuedness somehow built into it. Similarly, if there were objective principles of right and wrong, any wrong (possible) course of action would have not-to-be-doneness somehow built into it. Or we should have something like Clarke's necessary relations of fitness between situations and actions, so that a situation would have a demand for such-and-such an action somehow built into it (Mackie 1977:40).

As pointed out by both Olson (2014:103 ff.) and Joyce (2016a:§4.1), this passage describes (at least) two distinct features of objective values. It is one thing to claim that (knowledge of) something *motivates* or *makes* agents perform some action, another to claim that it *tells* or *demands* agents to perform it. Something can have one feature without having the other. For this reason, claims Olson (2014:105), this passage points to two arguments for error theory that Mackie could make: one focusing on the *motivating force* of moral facts, the other focusing on *the normative force* of moral facts. The latter argument contains traces of the argument from irreducibility. Olson says that

¹⁰ Because the traces of the argument from irreducibility are most clear in the *metaphysical part of the argument from queerness* (see n. 4), the other arguments are omitted from consideration here.

[...] talk of objective facts *telling* people to act so and so, and [...] *demanding* certain courses of action, is [...] akin to what is nowadays a widely accepted explication of the notion of a normative reason. According to this explication, a normative reason for an action is a fact that *counts in favour* of that action. The *favouring* relation is normative, as are the *telling*-relation and the *demanding*-relation (Olson 2014:105).

We will see in the next chapter that the argument from irreducibility focuses on *counting in favor-relations*. According to this argument, morality involves an error because ordinary moral thought and discourse involves a commitment to *irreducibly normative* counting in favor-relations (Olson 2016a:40-1, 2016b:463).

This section has given an account of Mackie's error theory. This account illustrated that Mackie accepts ERROR THEORY. The notions of a categorical imperative and a categorical reason, central to Mackie's ontological claim, have been contrasted to those of a hypothetical imperative and a hypothetical reason. Towards the end of the section, the possibility of interpreting the arguments from non-institutionality and irreducibility as elaborations of arguments indicated by Mackie was explored. The remainder of this chapter discusses the argument from non-institutionality. The argument from irreducibility is discussed in the next chapter.

3.2 The argument from non-institutionality

This section discusses the argument from non-institutionality. According to this argument, which is suggested by Joyce (and possibly by Mackie), morality involves an error because moral discourse involves the false assumption that there can be *non-institutional categorical requirements* on agents to do something, so that agents can have *non-institutional desire-transcendent reasons* for action (Joyce 2011a:523). Because this assumption has been incorporated into the meaning of moral terms, all (positive) moral statements are uniformly untrue

I proceed as follows: First, I give an account of the argument from non-institutionality and its central notion of an institutional requirement. Second, I give an account of Joyce's argument for the semantic claim of the argument, and suggest that his claims about institutionality and non-institutionality can be understood in terms of *authority*. Third, I raise two challenges to the semantic claim.

The argument from non-institutionality¹¹ says that:

¹¹ Joyce (2011a:523) refers to this argument as the Mackie/Joyce argument for error theory.

- (J₁) [...] moral discourse presupposes non-institutional desire-transcendent reasons and non-institutional categorical imperatives (Joyce 2011a:523).
- (J₂) [...] all genuine desire-transcendent reasons are institutional and all genuine categorical imperatives are institutional (ibid).¹²

If (J₁) and (J₂) are both true, we have good reasons to believe that moral discourse involves an error. But ERROR THEORY is true only if no positive moral proposition is true (i.e., only if FAILURE is true). Consequently, (J₁) and (J₂) amount to ERROR THEORY only if (J₁) is interpreted as the semantic claim that moral discourse *presupposes* non-institutional desire-transcendent reasons and non-institutional categorical requirements *specifically in such a way that* a positive moral proposition is true only if there are such reasons and requirements. When (J₁) is interpreted in this way, (J₁) and (J₂) entail, first, that a moral statement that an agent, *S*, (morally) ought to φ is true only if there is a non-institutional categorical requirement on *S* to φ and *S* (therefore) has a non-institutional desire-transcendent reason to φ , and, second, that no such statement is true.

To evaluate (J₁) and (J₂) for truth, we need to know what an *institutional requirement* is. In Mackie's sense of the term, an *institution* is a persisting and maintained social practice, "constituted by many people behaving in fairly regular ways, with relations between them which transmit and encourage and perhaps enforce those ways of behaving (Mackie 1977:80)."

Examples of institutions include chess and promising (ibid). A promise, it seems, can only be made in the presence of what we can refer to as "the promising institution". Suppose that Tony wants to promise Jessica that he will φ . In order for him to do so, it is not sufficient that he utters "I promise you that I will φ " in the presence of Jessica while intending to φ , that Jessica desires that φ and forms the expectation that Tony will φ , etc. For Tony's behavior to amount to a promise to Jessica that he will φ , and mean that he has taken on an obligation to φ , his behavior has to be "embedded in and reinforced by general social expectations, approvals, disapprovals, and demands (ibid:81)."

An institution has rules or principles of behavior (or both), and a set of concepts associated with it. The rules and concepts are interdependent: the concepts can be fully explained only in relation to the rules, and the rules are usually formulated partly in terms of these very concepts (ibid:80). The rules associated with an institution are formulated, encouraged, and enforced by its participants. By taking the rules as guide to their behavior, encouraging conformity with the rules, and discouraging violations of them, participants make demands both on themselves and others to act according to the rules.

¹² Garner (1990:143) makes a claim similar to (j₂).

When participants assert or invoke the rules, they use the institution's associated concepts. By speaking within the institution in this way – rather than describing it from the outside – participants *help constitute* the institution (ibid:81). The institution, recall, *is constituted by* regular behavior and discourse that transmits, encourages, and enforces this very behavior and ways of speaking.

Requirements belonging to an institution – *institutional requirements* – are “constituted by human thought, behaviour, feelings, and attitudes (ibid).” Suppose that Miranda watches Brad playing a game of chess, and expresses a requirement not to move the king more than one square, directed at him. There are two kinds of requirements in play here: the chess requirement not to move the king more than one square per turn, and the requirement Miranda makes on Brad *by invoking* the chess requirement. It is clear that Miranda's requirement is mind-dependent in that it is constituted by her thoughts, behavior, feelings, and so on. Consequently, her requirement is *subjective* in this sense. The chess requirement, however, is not constituted by Miranda's thoughts, behavior or feelings. Still, the claim that institutional requirements are constituted by human thought, behavior, feelings and attitudes, implies that the chess requirement is *also* mind-dependent in a central sense, and hence also subjective (Joyce 2016a:§4.1). This suggests that a requirement that was neither an agent's requirement, nor an institutional requirement would be *objective* in the sense of being *mind-independent*. It would be a requirement made without reference to anything of our own creation (Joyce and Kirchin 2010:xix).¹³ Given this, it is reasonable to suppose that when Mackie objects to *objectively valid* categorical imperatives, he objects to (categorical) requirements that are neither agents' requirements, nor institutional requirements.¹⁴

We are now in a position to see that the following passages from *Ethics* speak in favor of the suggestion that Mackie thinks that all *genuine* moral reasons and requirements are institutional (Joyce 2011a:523):

Do the desires and especially the sufferings of other people, if known to me, constitute a reason for me to do something, if I can [...]? It would be natural to say that they constitute some reason [...]. But the important thing is that if we recognize this as a further class of reasons, independent of any desire that I now have to help these other people, we are [...] bringing in the requirements of something like an institution [...] (Mackie 1977:78-9).

¹³ Note that it is possible for an agent, S_1 , to categorically demand something of another agent, S_2 , without invoking a requirement belonging to some institution (including morality). For example: If S_2 is standing on S_1 's foot, S_1 might categorically demand that S_2 moves.

¹⁴ For a different view on what Mackie means by “objectivity”, see Miller (2013:108-9).

[...] we may well say, ‘Surely if someone is writhing in agony before your eyes [...] this is in itself, quite apart from your feelings, a reason for you to do something about it if you can; if you don’t admit that, you just don’t know what a reason is, you can’t be using the word “reason” with its ordinary meaning, you can’t have the full concept of a reason.’ But if we say this, we are [...] speaking within the institution (ibid:79).

The argument from non-institutionality, recall, says that

- (J₁) [...] moral discourse presupposes non-institutional desire-transcendent reasons and non-institutional categorical imperatives (Joyce 2011a:523).
- (J₂) [...] all genuine desire-transcendent reasons are institutional and all genuine categorical imperatives are institutional (ibid).

Suppose that Mackie is right that all genuine moral (desire-transcendent) reasons and all genuine (categorical) moral requirements are institutional. If so, (J₂) *as applied specifically to moral reasons and requirements* is true, and the success of the argument from non-institutionality turns on whether there are any good reasons to believe (J₁).

Joyce’s argument for the semantic claim

I claimed above that the argument from non-institutionality entails ERROR THEORY only if (J₁) is interpreted as the semantic claim that moral discourse presupposes non-institutional desire-transcendent reasons and non-institutional categorical requirements *specifically in such a way that* a positive moral proposition is true only if there are such reasons and requirements.

Joyce defends (J₁) by arguing that the suggestion that moral reasons and requirements are *institutional* goes against the way we think of morality. According to him, we are not content to admit that our claim that there is a reason to refrain from killing is merely a permissible way of speaking from a perspective that endorses the dictates of morality (Joyce 2001:36-7, 40-2, 2006:63, 202). One way in which Joyce expresses this claim is as follows:

The rules and reasons of [...] [a] game [...] may quite reasonably be entirely ignored. We do not, I reiterate, think of morality in this way. Morality purports to have more authority than this – it is not something that a person may escape (in David Wiggins’s words) “by simply flying the skull and cross-bones and renouncing altogether the aim of belonging to the moral community.” (Joyce 2001:41-2)

Joyce elaborates on the difference in how we think of morality and normative institutions (like chess and etiquette) by suggesting that while *moral* reasons are “real reasons (ibid:40)” that we “cannot legitimately ignore (ibid)”, *institutional* reasons are “non-binding (ibid:41)”, and “may quite reasonably be entirely ignored (ibid).” According to him, we intuitively think “that a person is bound by [...] [the moral] rules whether he accepts them or not (ibid:34)” and that these rules are in some sense *his* rules even if he explicitly rejects them (ibid). Joyce also claims that if morality were an institution, people who genuinely don’t care about it would be “legitimately free to ignore it (Joyce 2006:63).”

This is not very elucidating. First, it is unclear how these claims should be interpreted. What is it, for example, for a requirement or reason to be “binding” and “real”, or such that it “cannot legitimately be ignored”? Second, it is unclear whether these claims describe merely one or several (perhaps related) differences between morality and institutions. While the claim that moral reasons cannot *legitimately* be ignored sounds like a normative claim, the claim that they are real sounds like an ontological or metaphysical claim.

A possible explanation as to why Joyce is vague on these matters is that his aim is to induce an *intuition* that there is a fundamental difference between morality and institutions, rather than to give an account of what this difference consists in.¹⁵ Joyce uses Plato’s example of Gyges to, as he puts it, motivate a sense of unease in his readers in thinking that morality might be an institution, and pump the intuition that our unease with this suggestion leaves us groping for non-institutional categorical imperatives and non-institutional desire-transcendent reasons.¹⁶ Consider, for example, the following passage:

Because our moral framework is categorical we can [...] legitimately [...] [say] “Gyges, you ought not to do that!” But if our utterances are merely a verbal output that has been validated by an institution of our own creation, then it all begins to sound rather shrill. We can picture Gyges pausing in some sadistic undertaking, acknowledging to us that, yes, we are speaking legitimately when we assert that he morally ought not to act in this way, and then shrugging and carrying on. And if our categorical imperatives are nothing more than an institutionally-backed way of speaking, why should he do otherwise? (Joyce 2011a:524)

¹⁵ Perhaps, this is all he *can* do. Joyce says: “I have come to suspect that my inadequacy in articulating this idea [i.e. non-institutionality] is not because of any failure of imagination or eloquence. Rather, morality may be imbued with a deeply mysterious kind of force [...] that resists explication. [...] Perhaps Mackie and I fumble to dissect something that by its very nature cannot be brought into the light to be picked over by philosophical scrutiny (Joyce 2011a:524-5).” See also Joyce (2007:60-61).

¹⁶ In Book II of Plato’s *The Republic*, Glaucon tells the story of a shepherd – Gyges – who finds a ring of invisibility which he uses to rape, kill and make himself king. See Plato (1997:1000-1001).

Joyce claims that if we catch Gyges in some sadistic undertaking, we want to be able to say something that constitutes a reasonable argument *for him* (ibid), and that

[w]e want to say both that Gyges ought not kill innocent people and that he has a reason not to kill innocent people. But we want this to be a “real” reason – a reason that somehow engages the shepherd – one that he cannot legitimately ignore. In other words, in the moral case, we are not content to admit that our claim that there is a reason to refrain from killing is merely a permissible way of speaking from a perspective that endorses the dictates of morality. We are still left with a desire to say something more – to imbue the moral imperative with a greater authoritative force – and we are still left with the unsettling possibility that it cannot be done (Joyce 2001:41).

According to Joyce, it is this *desire to say something more* that leaves us groping for non-institutional requirements and reasons (Joyce 2001:41, 2011a:523-4).

One way to make sense of Joyce’s claims about institutionality is in terms of (practical) *authority*. I will say that

- (B₁) A rule or norm, *N*, has authority for an agent, *S*, *if and only if* *S* takes *N* as authoritative
- (B₂) *S* takes *N* as authoritative *if and only if* *S* takes *N* as providing reasons relevant for practical deliberation
- (B₃) *S* takes an *institutional* norm or rule, *N*, as providing reasons relevant for practical deliberation *only if* *S* endorses the institution with which *N* is associated.

When taken together, (B₁) and (B₂) imply that a norm has authority for an agent *if and only if* she takes it to provide reasons relevant for practical deliberation. Together with (B₃), this in turn implies that an institutional norm has authority for an agent only if she endorses the institution with which it is associated. Endorsing an institution essentially involves endorsing those rules or principles of behavior which are associated with it. In the case of the *moral* institution, endorsing it essentially involves endorsing *moral* norms.

Let us start by considering Joyce’s claim that if morality were an institution, people who genuinely don’t care about it would be legitimately free to ignore it. Given (B₁)-(B₃), this claim can be interpreted as follows: if morality were an institution, moral norms would have authority only for those who cared

about morality enough to choose to endorse the moral institution. Consider, for illustration, etiquette. Because we are not logically forced to participate in this institution (Mackie 1977:79-80), and people have different attitudes towards etiquette, it is reasonable to suppose that only those who care about etiquette (to some sufficient degree) will endorse it. Suppose that Chad genuinely does not care about etiquette. He may still *know* what the contents of the etiquette rules are and that requirements of etiquette are often expressed as categorical imperatives. Chad may even consistently admit (e.g.) that according to etiquette, there is a categorical requirement on him not to eat soup with the spoon in his left hand. But because he does not care about etiquette, his recognition that these rules require this of him has no effect on him when he reflects on what to do. Because he does not take the rules of etiquette as authoritative, he may, *consistent with the standards he accepts as authoritative* wholly ignore these rules in practical deliberation. Consequently, we may say that because Chad genuinely does not care about etiquette, he may *legitimately* ignore it. Analogously, if morality were an institution, people who genuinely do not care about it would be legitimately free to ignore it.

Second, consider Joyce's claim that we want Gyges to have a "real" reason not to kill innocent people – a reason that somehow engages him and that he cannot legitimately ignore. Given (B₁)-(B₃), this claim can be interpreted as follows: We want moral norms to have authority for Gyges – i.e., we want him to take moral norms as providing reasons relevant for practical deliberation. If morality is an institution, moral norms have authority for Gyges only if he cares about morality to some sufficient degree. Because we have strong reasons to believe that Gyges genuinely does not care about morality at all, the fact that we nonetheless intuitively *want* moral norms to have authority for him – if we do indeed want this – can be taken to support the claim that we intuitively *do not* think of morality as an institution (or at least that we intuitively *do not want* it to be an institution).

Consider, lastly, the argument from non-institutionality. I claimed above that while the notions of "institutionality" and "non-institutionality" are central to this argument, it is not entirely clear how Joyce intends for these notions to be understood. I introduced the notion of "authority" as a way of making sense of them. The second premise says that all genuine desire-transcendent reasons and categorical imperatives are institutional. Given (B₁)-(B₃), this claim can be interpreted as implying the following: all genuine moral reasons and requirements have *institutional authority* – i.e., are taken to provide reasons relevant for practical deliberation only by agents who endorse the moral institution (because morality *is* actually an institution, moral reasons and requirements are institutional reasons and requirements). The first premise of the argument says that moral discourse presupposes non-institutional desire-transcendent moral reasons and non-institutional categorical moral requirements specifically in such a way that a positive moral proposition is true

only if there are such reasons and requirements. Given (B₁)-(B₃), this claim can be interpreted as follows: a positive moral proposition is true only if there are moral reasons and requirements that have *non-institutional authority* (whatever such authority consists in).¹⁷ Consequently: given (B₁)-(B₃), the argument from non-institutionality can be interpreted as saying that morality involves an error because a positive moral proposition is true only if moral reasons and requirements have *non-institutional authority*, and all *genuine* moral reasons and requirements have *institutional authority*.

Moral practice as an institution

This and the following sub-section challenges the argument from non-institutionality by raising two challenges to (J₁) – i.e., the semantic claim that moral discourse presupposes non-institutional desire-transcendent reasons and categorical requirements specifically in such a way that a positive moral proposition is true only if there are such reasons and requirements.

Because I find it difficult to imagine how there could be categorical requirements that were not the requirements of agents or institutions (or that had non-institutional authority), I grant Joyce and Mackie (J₂) as applied specifically to moral reasons and requirements – i.e., the claim that all genuine desire-transcendent moral reasons and all genuine categorical moral requirements are institutional.¹⁸

In this sub-section, I challenge (J₁) by challenging the claim that moral discourse presupposes specifically *non-institutional* reasons and requirements. I will do so by exploring the possibility of understanding moral discourse as presupposing merely *institutional* reasons and requirements – reasons and requirements that have authority only for those who endorse the moral institution.¹⁹ In the next sub-section, I challenge (J₁) by challenging the claim that moral discourse presupposes non-institutional requirements and reasons

¹⁷ Given that *institutional* authority is a matter of being taken, by an agent, to provide reasons relevant for practical deliberation only if she endorses the relevant institution, and that an agent will endorse an institution only if she cares about the institution and/or its associated norms and rules to some sufficient degree, one might be tempted to understand *non-institutional* authority as a matter of being taken to provide reasons relevant for practical deliberation by an agent *regardless of* whether (or to what degree) she cares about the relevant institution and/or its associated norms and rules. But if Joyce suggests that moral discourse involves the assumption that *morality is not an institution* – as he seems to do – he would probably not describe moral discourse as involving the assumption that moral reasons and requirements are taken, by an agent, to provide reasons relevant for practical deliberation regardless of whether, or to what degree, she endorses the *moral institution*.

¹⁸ In other words, I will use what Joyce (2011a:519-20, 2013:7) terms a “concessive” strategy: granting that we should not believe in non-institutional categorical requirements, while insisting that moral discourse need not be understood as committed the claim that there are such requirements, or at least not as committed to it in such a way that moral discourse involves a fundamental and serious error. (In contrast, to use a “head-on” strategy is to offer arguments in support of the claim that we should believe in non-institutional categorical requirements.)

¹⁹ Thanks to Gunnar Björnsson for bringing the possibility of interpreting morality as an institution to my attention.

specifically *in a semantic sense* which entails that a positive moral proposition is *true* only if there are non-institutional requirements and reasons.

Joyce claims that the suggestion that morality is an institution goes against the way we think of it. But this can be questioned. It seems to me that morality can be understood as an institution, and that those features of moral thinking and discourse which Joyce takes to support the claim that we *do not* think of morality as an institution can be explained also on the assumption that we *do* think of it as such. To the extent that an institutional account can explain features of moral thinking and discourse that we take to be peculiar to it, the need for a non-institutional account diminishes.

If morality is an institution, it is analogous to other normative institutions like etiquette and chess in the sense that moral norms have authority only for those who endorse it. Agents who genuinely do not care about morality will not take moral norms as providing reasons relevant for practical deliberation. We saw above that Joyce takes the fact – if it is a fact – that we want Gyges to have a “real” reason not to kill innocent people to speak in favor of the claim that the suggestion that morality is an institution goes against the way we think of it. But it seems to me that the fact that we *want* this is compatible with morality’s being an institution, and our thinking of it as such. The reason for this is that Gyges is a rare exception. First, because most of us care deeply about morality, most of us do endorse the moral institution (including the moral norms), and are strongly committed participants in it. Most of us do take moral norms as providing reasons relevant for practical deliberation that we cannot, by our own standards, legitimately ignore. Second, because we are aware of the fact that most of us are alike in this respect, we normally take it for granted that others care about morality and take moral norms as authoritative. Third, because we care about morality, we find it important that moral norms are taken as authoritative, both by ourselves and others. For these reasons, we normally both want and expect other agents we encounter – even Gyges – to accept moral norms as authoritative.

If morality is an institution, it is analogous to other institutions also in the sense that a distinction can be made between how we think, speak, and reason from within it (as participants) and from outside it (as observers). We saw above that Joyce claims that when we make moral statements, we are not content to admit that we are simply speaking in a way that is permissible from a perspective that endorses morality. We also saw that Joyce takes this fact – if it is a fact – to speak in favor of the claim that we do not think of morality as an institution. Mackie similarly claims that when we use the concept of objective moral value, we are supposing that there are requirements which simply are there, without being the requirements of any person or body of persons, even God (Mackie 1977:59). But it seems to me that both these claims are compatible with the suggestion that we think of morality as an institution, because it can be argued that they describe how we normally proceed *when we are caught up in*

the moral institution – i.e., how we normally think, speak and reason from within the moral institution. Suppose that it is true that when an agent makes a moral judgment or statement, she sometimes thinks or says to herself that moral requirements are neither agents' requirements nor institutional requirements, and that she is not simply speaking in a way that is permissible from a perspective that endorses morality. It seems to me that even if she makes such assumptions in contexts where she makes moral judgments or statements, this is consistent with the possibility of her being aware – even when caught up in the moral institution – that moral norms have authority only for those who endorse it, and that regardless of how much she wants Gyges to have a real reason not to kill the innocent, there is no sense to be made of the claim that he does.²⁰

If morality is an institution, it differs from other normative institutions in *scope* and *degree*. Morality has greater scope than etiquette and chess in the sense that moral norms apply to agents located in a wide set of loosely defined contexts (i.e. those where one's behavior affects others or has the potential to do so), and that in any context where etiquette norms or chess rules apply (e.g. contexts where one plays a game of chess, or participates in some formal social context like a wedding dinner), moral norms apply too. Moreover, a violation of etiquette or chess is often also a moral violation. For example, it can be argued that under normal circumstances, to cheat in a game of chess is to break a (tacit) contract with one's opponent. Something similar might be true about etiquette violations. It can be argued that under normal circumstances, to participate in a wedding dinner while violating the social codes for formal social contexts of this type is to harm other participants in this context.

The difference in degree between morality and other institutions is twofold. First, our attitudes (broadly understood) to moral reasons and requirements differ in strength from our attitudes to the reasons and requirements of other institutions. Most of us *care* about morality and its associated reasons and requirements. And although we care also about the reasons and requirements of other institutions, we generally care more about those of morality. For example, we generally find it more important that people (ourselves and others) act according to the moral rules we endorse than that they act according to rules of etiquette or chess we endorse; we are generally more eager to impose our moral norms onto people whose moral norms differs from ours than we are to impose etiquette or chess rules onto those whose rules of etiquette or chess differ from ours; we generally react more strongly to moral violations than etiquette or chess violations. Moreover, given that we care more about morality than etiquette and that in any context where etiquette norms apply, moral norms apply too, it may be argued that to the extent that we care

²⁰ The situation is analogous to that when we are caught up with a fiction. In chapter 8, I argue that it is possible for agents to be aware that they are engaged with a fiction and still be motivated by this engagement.

about (e.g.) whether someone wears a red dress at a wedding dinner, this can partially be explained with reference to the fact that we care about the moral norm according to which one must avoid inflicting intentional harm on others.²¹

The second difference in degree between morality and other institutions is that we generally give moral considerations precedence in practical inter- and intra-personal deliberation. When reflecting on what someone ought to do in some situation, we generally treat moral considerations as holding greater weight than non-moral considerations, including considerations of etiquette or chess. This difference is related to the first in that it is reasonable to suppose that the reason why moral considerations take such precedence is precisely that we *care* about them, and that we care *more* about them than non-moral considerations.

Taken together, the claims above – i.e., that most of us endorse the moral institution and take its norms as authoritative, and that morality has greater scope than, and trumps, other institutions – suggest that we normally proceed as if moral norms have authority, and normally think, speak and reason from a perspective internal to morality. Because we take moral considerations to have precedence over other institutions in practical deliberation, we operate from a perspective internal to morality not only in explicitly moral contexts – e.g., when trying to solve moral dilemmas or to reach agreement on some set of normative principles – but in a wide variety of contexts. All of this is consistent with the suggestion that morality is an institution, and that we think of it as such.

Moral authority as a shared background assumption

The first premise, (J₁), of the argument from non-institutionality is, again, the semantic claim that moral discourse presupposes non-institutional desire-transcendent reasons and non-institutional categorical requirements specifically in such a way that a positive moral proposition is true only if there are such reasons and requirements. In the previous sub-section, I challenged (J₁) by exploring the possibility of understanding moral discourse as presupposing merely institutional reasons and requirements. In this sub-section, I raise a second challenge to (J₁). More specifically, I challenge the claim that moral discourse presupposes non-institutional requirements and reasons specifically *in a semantic sense* which entails that a positive moral statement is true only if there are non-institutional requirements and reasons. I offer an analysis of moral discourse which illustrates that, contrary to what Joyce and Mackie suggest, it is possible that the relevant assumption enters pragmatically into moral conversations, and does not affect the truth values of moral

²¹ The claim made here concerns people in general. For this reason, it is compatible with the possibility that *some* people – some highly dedicated chess players, say – are more concerned with chess rule violations than almost anything else.

judgments and statements. This analysis connects to the discussion in the previous sub-section in the following way: the claim that most of us endorse the moral institution and take its norms as authoritative, and that we are aware of this, is congenial to the claim that the relevant assumptions enter pragmatically into moral conversations.

The argument from non-institutionality says that while all genuine categorical requirements and all genuine desire-transcendent reasons are institutional, moral discourse presupposes *non-institutional* such requirements and reasons. This entails ERROR THEORY only if a moral statement like

- (e) Brad ought to give to famine relief,

semantically implies a statement the truth conditions of which obtain only if there is a non-institutional categorical moral requirement on Brad to give to famine relief, so that Brad has a non-institutional desire-transcendent moral reason to do so.²² But the claim that moral statements imply such statements can be questioned. It can be argued that even if we believe or presuppose that moral requirements provide agents with non-institutional desire-transcendent reasons, this need not affect the semantics of moral language.²³

It seems to me that it is a shared background assumption in moral conversations that moral norms *have (practical) authority* (i.e., moral norms provide reasons relevant for practical deliberation), that moral considerations *take precedence in practical deliberation* (i.e., moral considerations have more weight in practical inter- and intra-personal deliberation than non-moral considerations), and *are taken as such by agents* (i.e., moral agents take moral norms as authoritative and give moral considerations precedence in practical deliberation). I will refer to the assumption that moral considerations have authority, take precedence in practical deliberation, and are taken as such by agents as “the authority assumption.” If the authority assumption is a shared background assumption in moral conversations rather than semantically encoded, it does not affect the truth values of moral statements. If so, the fact – if it is a fact – that participants in moral conversations assume that considerations have a kind of authority which they actually do not have, or which cannot be made sense of, does not imply that there are no true positive moral statements. Consequently, the account of moral discourse given below is compatible with the truth of positive moral statements, and therefore with the falsity of ERROR THEORY.

²² Joyce seems to hold that moral statements entail statements about non-institutional categorical requirements because it is necessary for the applicability of moral concepts – i.e. concepts like *morally wrong* or *morally ought (to be done)* – that there exist such requirements (Kalf 2013:925).

²³ Daan Evers (2016:408-9) calls the question whether our ways of thinking about the nature of moral reasons affect the semantics of first-order moral language “the million dollar question” for moral error theory.

The suggestion that the authority assumption is a shared background assumption in moral conversations is, more specifically, the suggestion that this assumption is a *pragmatic presupposition* when moral sentences (i.e., unembedded moral sentences in the declarative form) are used. Normally when someone makes a moral statement, she takes it for granted that (i) moral requirements and reasons have practical weight and take precedence in practical intra- and interpersonal deliberation and reasoning, that (ii) her addressees take it for granted that moral considerations have such weight and take such precedence, and that (iii) her addressees recognize that she is taking this for granted. (I take this account of pragmatic presupposition to be consistent with that provided by Robert C. Stalnaker 1974:200, 202.²⁴)

A pragmatic presupposition associated with sentence, *s*, is a condition, *x*, such that when a speaker or uses *s*, she normally expects *x* to hold as common ground between herself and her audience (Beaver and Geurts 2014:§4.2). It is in virtue of the fact that normally when someone uses *s*, she expects *x* to hold, that *x* is associated with *s*. When a speaker uses

(f) Chad doesn't know that the apples are ripe,

to make a statement, she normally takes it for granted, and her audience will normally interpret her as taking it for granted, that

(g) The apples are ripe.

But it is also possible to use (f) to make a statement while taking for granted that

The apples are *not* ripe.

A speaker might use (f) in a context where she wants to communicate that because the apples are in fact not ripe, Chad doesn't *know* that the apples are ripe, but falsely believes that this is so. Consequently: While (f) is associated with (g) in the sense that *normally* when someone uses (f), she takes (g) for granted and is interpreted as such, using (f) is compatible also with taking the denial of (g) for granted. While (g) is a pragmatic presupposition associated with (f), (f) does not semantically imply (g).²⁵

It is important to note that it is agents – speakers and audiences – that make or have presuppositions, not sentences, propositions, or other linguistic entities (Stalnaker 1974:200). Facts about what a speaker (and her audience)

²⁴ Stalnaker gives the following account: “[...] proposition P is a pragmatic presupposition of a speaker in a given context just in case the speaker assumes or believes that P, assumes or believes that his addressee assumes or believes that P, and recognizes that he is making these assumptions, or has these beliefs (Stalnaker 1974:200).”

²⁵ This example is a modified version of an example given by David I. Beaver and Bart Geurts (2014:§4.2).

normally takes for granted when she uses a certain sentence to make a statement are distinct from semantic facts about the sentence used. For this reason, it is possible for the same semantic content – indeed, for the very same unambiguous *sentence* – to be associated with different pragmatic presuppositions depending on context, intonation, or word order (ibid:204).

Presuppositions can be triggered by various things. According to my suggestion, the relation between moral sentences like

- (h) Miranda (morally) must not kill innocent people,

and some assumption about the authority of morality, for example

moral requirements, such as the requirement on Miranda not to kill innocent people, provide reasons relevant for practical deliberation

is of the same kind as the relation between the sentence (f) and the assumption described in (g), above. In both cases, speakers using the first sentence normally take for granted, and are normally interpreted as taking for granted, the assumption described in the second sentence (or something that resembles it). Moreover: given COGNITIVISM, these cases are analogous in the sense that normally when a speaker uses the first sentence in a conversation, she is thereby making an assertion.²⁶

An important difference between these cases, however, is that the presuppositions are brought to our attention (“triggered”) by different things. When someone uses (f), the presupposition is triggered by the fact that the sentence used involves the term “know”. Speakers using the term “know” typically presuppose that its complement (in this case, that *the apples are ripe*) is true (Beaver and Geurts 2014:§4.2). However: when someone uses (h), the presupposition is triggered not merely by the terms involved in this sentence, but also by something else. I suggest that this other thing is a feature of contexts where moral sentences are typically used (see Kalf 2013:931 for discussion). The set of contexts where moral sentences are typically used includes, but is not limited to, moral conversations. The relevant contexts are contexts where the question at issue is understood to be *practical*. What makes something such a context might vary, but it will partly be guided by the default assumption that people care about morality, and also by the default assumption that what makes moral questions worth discussing is precisely that we take answers to them to be

²⁶ I said in chapter 2 that according to COGNITIVISM, unembedded moral sentences in the declarative form are typically used to make moral assertions. A moral assertion is a speech act in which a moral proposition is claimed to hold. I also said that the context in which an utterance is made has relevance for whether it counts as an assertion. (I return to this issue in chapters 7 and 8.) Because (h) is an unembedded moral sentence in the declarative form, COGNITIVISM implies that (h) is conventionally used to make an assertion.

relevant for what to do and how to feel. The words “ought” or “should” will, in many contexts, trigger the assumption that we are concerned with a moral question of practical importance, but this will depend subtly on context, since these words are also used to express others sorts of practical modalities, as well as epistemic modality.²⁷

For an illustration of my suggestion that the authority assumption is a shared background assumption in moral conversation, suppose that our mutual friend, Mark, got drunk and cheated on his girlfriend, Stephanie, last weekend, and that you and I are discussing how to help him out of his current predicament (and ultimately what action we should recommend that he take). Suppose that you utter

- (i) Mark ought to tell Stephanie the truth,

invoking recognizably moral considerations in support. Supposing that our conversation takes place under normal conditions, COGNITIVISM implies that by uttering this moral sentence, you have made an assertion, and the claim the authority assumption enters into moral conversations as pragmatic presuppositions of speakers – as I intend for it to be understood – has two implications. First, it implies that you and I both take something like the following for granted about the authority of moral requirements and reasons:

moral considerations are to be given more weight in our deliberation than non-moral considerations.

Second, it implies that you and I both take something like the following for granted about each other (our respective addressees):

my addressee takes it for granted that moral considerations are to be given more weight in our deliberation than non-moral considerations, and recognizes that I am also taking this for granted.

It seems reasonable to suppose that conversations are normally cooperative efforts such that participants in any conversation normally recognize a common aim (or mutually accepted direction) of the conversation, and intend for their utterances to be relevant contributions to this aim.²⁸ Given this, it is reasonable to suppose that when a speaker makes a moral assertion, she normally takes it for granted that the moral consideration thereby brought to attention *has relevance* for the issue presently discussed. Given that the aim of our conversation, in the present example, is to figure out what action to recommend

²⁷ Thanks to Gunnar Björnsson for helpful discussion here.

²⁸ For more on the idea that conversations are cooperative efforts, see Grice (1975:45 ff.).

to Mark (and that conditions are otherwise normal), this suggests that when you utter (j), you take for granted, and I take for granted that you take for granted, something like the following:

if Mark morally ought to tell Stephanie the truth, this would be relevant for determining what advice we should give Mark.

In other words: it seems reasonable to suppose that when you use the moral sentence (j) in the context of the present example, you and I both take for granted that the moral consideration thereby brought to attention has practical relevance. I suggested, above, that for a norm or rule to have (practical) authority is for it to be taken by agents as providing reasons relevant for practical deliberation. There is a clear connection between the assumption that some moral consideration has practical relevance and the assumption that it has (practical) authority. To assume, without objection, that a moral consideration has practical relevance is to accord it authority.

Some consequences of my suggestion

Although some of the above discussion was made under the assumption that moral sentences are used in the way COGNITIVISM suggests, the suggestion that the authority assumption is a shared background assumption in moral conversations should not be understood as depending on this claim. Even if moral sentences are not (typically) used to make assertions, it might be the case that when we utter moral sentences in contexts of moral conversations (and other contexts where the question at issue is understood to be practical), we normally take the authority assumption for granted, and are normally interpreted as doing so. And even someone who does not accept COGNITIVISM may agree that we frequently utter moral sentences in such contexts.

Nor should the claim that the authority assumption enters into moral conversations as speaker's presuppositions be understood as the suggestion that all agents engaging in moral discourse have the same conceptions of, and intuitions about, the ways in which moral norms and considerations are authoritative. To the extent that people's desires, goals, cultural heritage, conceptions of the world etc., vary, it is reasonable to expect that even people who care deeply about morality care about it in different ways, and have different conceptions of, and intuitions about, morality's authority. According to Joyce, recall, we think of moral rules as binding for agents whether they accept them or not, and of moral reasons as "real" reasons that cannot legitimately be ignored. One may or may not share these intuitions. (I am myself not convinced that I share each one of them.) My suggestion allows for such variations. First, it should merely be understood to concern what is *normally* taken for granted in moral conversations (and other contexts where the question at issue is

understood to be practical). Second, one and the same unambiguous moral sentence can be associated with different assumptions by different speakers, or by one and the same speaker in different contexts.

The fact that the assumptions associated with one and the same sentence can vary with speakers and contexts suggests that we *do not*, for any specific sentence, *s*, have to identify any specific assumption, *x*, such that in general, when a speaker uses *s* in the context of a moral conversation, she takes *x* for granted. It also suggests that we do not have to identify any specific assumption, *y*, such that in general, when speakers use moral sentences in moral conversations, they take *y* for granted. However, I do believe that we can come close to a *minimal* generalization of the latter kind, since I believe that we normally take the authority assumption for granted in such contexts. This generalization is ‘minimal’ in the sense that it can be combined with a variety of more specific assumptions about the way in which morality is authoritative. I believe, in other words, that agents normally take the authority assumption for granted when participating in moral conversations regardless of how they conceive of the further details of morality’s authority.

Because my suggestion allows for variation in what is taken for granted in moral conversations, it is compatible with it being the case that some speakers, in some such contexts, take for granted precisely those claims Joyce and Mackie take to be entailed by moral statements. In other words, it is compatible with it being the case that in the context of the conversation described above, you use

- (i) Mark ought to tell Stephanie the truth,

to make a moral statement while taking for granted some, or even all, of the following assumptions

- (j) The fact that Mark ought to tell Stephanie the truth is something *he cannot legitimately ignore*.
- (k) Mark has a *real reason* to tell Stephanie the truth.
- (l) Mark has a *non-institutional categorical (moral) reason* to tell Stephanie the truth.

At the same time, my account does not presuppose that the notions figuring in (j)-(l) have any clear meaning (whether cognitive or non-cognitive). For all my account says, it might be that they express a jumbled mess of thoughts and emotional or motivational dispositions.

One of the strengths of my suggestion lies in the fact that it is robust with respect to detailed accounts of the exact contents of the presuppositions of moral discourse. *Whatever* the contents of the problematic presuppositions of

moral discourse are, it is possible that these are merely pragmatic presuppositions of speakers engaging in moral conversations, rather than part of the content of moral statements and judgments. The account stands even if it turns out that the presuppositions of moral discourse lack definite content, but represent nothing more than a vague feeling of unease with various naturalistic or non-cognitivist analysis of moral judgments.

Another advantage of this suggestion is that the falsity of a pragmatic presupposition does not make the associated sentence semantically flawed (Beaver and Guerts 2014:§4.2). If a sentence, *s*, *pragmatically presupposes* an assumption, *x*, and *x* is false, this does not entail that *s* is false, nor that *s* cannot be evaluated for truth. Consequently, if it would turn out that there is in fact some false claim, *y*, such that when speakers make moral statements, they normally expect *y* to hold as common ground between themselves and their audiences, this would not be sufficient for it to be the case that no positive moral statement is true. In order for this to be the case, the presupposition somehow has to affect the propositional contents of moral statements.

Stephen Finlay formulates an argument congenial to the claim that the fact that a sentence, *s*, is associated with a false presupposition does not guarantee the falsity of *s*. I find this analogy, supplied by Finlay, illuminating:

For centuries, water was almost universally assumed to be an element rather than the compound it actually is. But we do not take seriously the [...] proposal that until Lavoisier, thought and talk about ‘water’ was systematically false because there was no such stuff—and we would not even if the assumption had been universal. Even if every person who engaged in moral discourse falsely assumed the absolute authority of moral value, therefore, it still doesn’t follow that their moral claims never succeed in being true (Finlay 2008:351-2).

James Drier also provides an example congenial to the claim that the fact that *s* is associated with a false presupposition does not in itself make *s* false. Dreier’s example runs as follows:

A policeman on the witness stand testifies that while staking out the apartment, he saw the defendant enter and then leave one hour later. The defense crossexamines: When you say it was one hour later, can you provide an inertial frame? “A duration of one hour must, officer, be relative to one inertial frame or another, you know.” The policeman denies that he meant any such thing. “Just one hour, is all I meant, not relative to any of your fancy frames” (Dreier 2006:261).

Like many of us, the policeman believes that *time is absolute*, although our best theory of physics tells us that *time is relative*. For this reason, the policeman takes something false for granted when he claims that he saw the suspect leave the apartment one hour later. But as Dreier points out

[...] the incorrect theory that the policeman himself would give if carefully questioned, does not seem to infect the integrity or veracity of his ordinary, first-order judgments. What the policeman said, we believe, is true; only his background absolutist theory of it is mistaken (ibid:262).

An additional reason why the fact – if it is a fact – that those assumptions we normally take for granted when we use moral sentences are false is not in itself sufficient for the truth of the claim that there are no true positive moral statements is that *pragmatic presuppositions can sometimes be cancelled*. Recall the example, above, where you and I were trying to figure out what advice to give our cheating friend Mark. I claimed that when you utter

- (i) Mark ought to tell Stephanie the truth,

you and I normally take for granted something like the following:

- (m) Moral considerations are to be given more weight in our deliberation than non-moral considerations.

But it is possible that you, for some reason or other do *not* consider the moral consideration brought to attention by (i) to take precedence in this particular argument. If so, you might want to indicate to me that although you recognize that there is a moral norm that applies to Mark – e.g. ‘one must be truthful to one’s spouse’ – you do not accept that moral considerations take precedence in this particular context. You want to use (i) without being interpreted as taking (m), or something similar to it, for granted. There are different ways in which you might achieve this. Consider, for example:

- (n) Mark morally ought to tell Stephanie the truth, but in the present argument, other kinds of facts take precedence.
- (o) Mark morally ought to tell Stephanie the truth. However, given that Mark wants to get a new girlfriend at some point in the future, I think that the best course of action for him is to focus on saving his reputation. If Stephanie finds out the truth, she will tell her friends about what he has done. News travel fast,

and if it becomes public that Mark is a cheater, he will never be able to get another girlfriend.

If you utter (i) as a part of (n) or (o), I will not interpret you as taking (m) or something similar for granted. If you utter (i) as part of (n), you indicate that you do not take (m) for granted by explicitly denying it. In (o), (m) is denied in a more implicit manner.²⁹ This feature of presuppositions (“cancellability”) is well recognized (see e.g. Kalf 2013:931 and Beaver and Geurts 2014:§3).

The fact that when you utter (n) or (o), you use (i) without taking it for granted that the moral consideration thereby brought to attention takes precedence in the present argument need not be taken to entail that (j) has a different *content* (or meaning) when used as part of (n) or (o), nor that (i) does not count as a *moral* sentence when used in such contexts. Moreover, the fact that you are not assuming that the moral consideration brought to attention by (i) takes precedence in the present argument does not entail that you are denying that moral considerations *normally* take precedence in practical interpersonal reasoning (which is what we are presently involved in). In fact, when you utter (n), it is natural to interpret you as indicating precisely that moral considerations normally take precedence in arguments like the present. Otherwise, your remark that in this specific case, there is a moral consideration (namely that Mark ought to tell Stephanie the truth) which does not take precedence would seem superfluous and unnecessary. Interpreting you as presupposing that moral considerations normally take precedence in practical interpersonal deliberation is more charitable than interpreting you as having made a statement that is not a relevant contribution to our aim of figuring out what course of action we should recommend that Mark take.

The fact that pragmatic presuppositions are cancellable seems to suggest that the claim that there are no true positive moral statements could not be saved by an argument against the claim that false background beliefs or assumptions do not necessarily make the associated statement false. Suppose that Joyce and Mackie would give convincing arguments for the claims (1) that if a moral statement is associated with a flawed pragmatic presupposition is not true, and (2) that there is a flawed assumption *x* such that normally when a speaker makes a moral statement, she presupposes *x*. If so they would not have shown that positive moral statements *used in contexts where the presupposition is cancelled* – e.g., (i), when used as part of (n) or (o) – are not true. For this reason, they would not have shown that *no* positive moral statement is true, and therefore would not show that ERROR THEORY is true.

In connection to this, it should be mentioned that there is a distinction to be made between the claim I have been defending – viz. that authority *need not*

²⁹ Similar moves can be made by a speaker claiming that Chad does not know that the apples are ripe in order to indicate that she is not taking for granted that the apples are ripe.

be a part of the meaning of moral statements – and the stronger claim that authority *is not* part of the meaning of moral statements. The combination of the claims that authority serves as a pragmatic presupposition in contexts where moral statements are made, and that pragmatic presuppositions are cancellable, could be used not only to defend the former claim, but also the latter. That is, it could be argued that the fact that the authority assumption can be cancelled shows that it is not part of the meaning of moral statements. However: since it is conceivable that cancellation does change the semantic content of the relevant ought-claim such that it no longer constitutes a moral claim proper, this conclusion does not follow directly from the possibility of cancelling the presupposition. Further argument would be needed to defend it.

A final consideration that can be addressed in support of the suggestion that the authority of morality is not a semantic, but rather a pragmatic, aspect of moral discourse is that the latter is sufficient for moral practice to fill its (supposed) functions – viz. for it to play its role in the regulation of actions, behavior, character traits (etc.), in order that such actions, behavior, character traits (etc.) that are harmful to civil society are generally avoided and that those that are favorable to civil society are promoted.³⁰ I believe that a case can be made that in order for moral practice to fill this function, it is sufficient that the authority of morality enters into moral conversations as pragmatic presuppositions. If this is the case, there is no need to complicate the semantics of moral discourse with the claim that moral statements (semantically) imply statements whose truth condition obtains only if there are non-institutional categorical requirements and non-institutional desire-transcendent reasons.³¹

In this sub-section, it has been suggested that the authority of morality is reflected in the pragmatics of conversations where moral sentences are used. Because this suggestion is consistent with the truth of COGNITIVISM and the falsity of FAILURE, it represents a way of accounting for the fact that we treat morality different than other normative systems while avoiding the conclusion that ERROR THEORY is true. Even granting both that all genuine moral requirements and reasons are institutional, and that moral discourse and thinking typically presupposes non-institutional requirements and reasons, it is possible that this does not entail ERROR THEORY, because it is possible that the non-institutional character of moral requirements and reasons is merely

³⁰ These issues are elaborated in chapter 7.

³¹ Motivational externalists like Sigrun Svavarsdóttir have argued that the appeal of motivational internalism can be explained by the fact that when we make moral statements, the most straightforward explanation of this is that we accord such statements practical significance. Although concerned with explaining why motivational internalism might seem true even though it isn't, rather than with explaining why the implication of authority might seem semantic even though it isn't, this explanation seems structurally similar to what I am proposing here, since both accounts invoke assumptions about the practical role of moral statements. For more on this, see Svavarsdóttir (1999).

pragmatically presupposed in moral discourse, rather than part of the content of moral judgments and statements.

3.3 Conclusion

This chapter has discussed J.L. Mackie's error theory and what has been referred to as *the argument from non-institutionality*, which has been suggested by Richard Joyce.

In the first section of the chapter, I gave an account of Mackie's error theory, which illustrated that Mackie accepts ERROR THEORY. I spelled out the notions of a categorical imperative and a categorical reason. I also explored the possibility of interpreting the arguments from non-institutionality and irreducibility as elaborations of arguments indicated by Mackie.

In the second section of the chapter, I challenged the argument from non-institutionality by raising two challenges to its first premise – i.e., the semantic claim that moral discourse presupposes non-institutional desire-transcendent moral reasons and non-institutional categorical moral requirements specifically in such a way that a positive moral proposition is true only if there are such reasons and requirements. First, I challenged the claim that moral discourse presupposes specifically *non-institutional* moral requirements and reasons. Even granting that all genuine moral reasons and requirements are institutional, it is possible to understand moral discourse and thinking as presupposing that moral requirements and reasons have authority only for those who endorse the moral institution. Second, I challenged the claim that the assumption that moral requirements and reasons are non-institutional enters *semantically* into moral conversations. Even granting both that all genuine moral requirements and reasons are institutional and that moral discourse and thinking typically presupposes non-institutional requirements and reasons, it is possible that the non-institutional character of moral requirements and reasons is merely *pragmatically* presupposed in moral discourse.

If these arguments are sound, there are reasons to doubt the first premise of the argument from non-institutionality. If so, the argument has not convincingly shown that moral thought and discourse involve an error of the kind that makes all positive moral judgments and statements untrue.

Chapter 4

The argument from irreducibility

This chapter discusses the argument from irreducibility. According to this argument, which is suggested by Jonas Olson (and possibly by J.L. Mackie), morality involves an error because moral thought and discourse involves a commitment to *irreducibly normative favoring relations*, and there are no such relations. My focus in this chapter will be on Olson's claim that moral claims (are or) entail claims about irreducibly normative reasons (Olson 2014:124). Olson defends this claim by responding to Finlay's (2006, 2008) relativistic account of moral discourse. I challenge the argument from irreducibility by challenging Olson's response to Finlay.

I proceed as follows: First, I give an account of the argument from irreducibility (the present section). Second, I spell out the notion of an irreducibly normative reason, central to the conceptual claim of the argument (section 4.1). Third, I challenge Olson's argument for the semantic claim that moral claims (are or) entail claims about irreducibly normative reasons. I give an account of Finlay's relativistic account of morality, and his challenge to "the disputation evidence" for absolutist (i.e., non-relativist) accounts of moral discourse. I argue that Olson's response to Finlay's challenge to the disputation evidence does not convincingly show that moral claims entail claims about irreducibly normative favoring relations (section 4.2).

The argument from irreducibility goes as follows:

- (O₁) [p] Moral facts entail that there are facts that favor certain courses of behavior, where the favoring relation is irreducibly normative.
- (O₂) [p] Irreducibly normative favoring relations are queer.
- (O₃) [1,2] Hence, moral facts entail queer relations.
- (O₄) [p] If moral facts entail queer relations, moral facts are queer.
- (O₅) [4] Hence, moral facts are queer (Olson 2014:123-4).

The (O₁)-(O₅) argument is what Olson calls *a queerness argument*. As such, it cannot in itself – i.e. without support from some good arguments for

projectivism or some debunking explanation of moral belief – show that moral error theory is more plausible than moral realism.¹ Olson (2016b:463) says that the queerness of moral facts is meant to establish only a presumption against the existence of moral facts, not a conclusive refutation. But it seems to me that premise (O₁) – i.e., claim that “[m]oral facts entail that there are facts that favor certain courses of behavior, where the favoring relation is irreducibly normative (Olson 2014:123)” – combined with the claim that “ordinary moral statements purport to refer to moral facts (Olson 2016a:397)”, and that “it is far from clear that there are, or can be, any irreducibly normative facts (ibid:401)” seems to suggest (although not strictly speaking imply) that ERROR THEORY is true.

For this reason, I believe that it is worthwhile discussing the argument from irreducibility without going into details of arguments for projectivism, debunking arguments, or the other three – in his view unsuccessful (see Olson 2014:3, 79, 2016a:400) – queerness arguments Olson examines.

4.1 The conceptual claim

Two central premises of the argument from irreducibility are (O₁) and (O₂). Olson (2014:124) refers to the claim expressed in (O₁) as “the conceptual claim.” More specifically, it says that

[...] moral facts are or entail irreducibly normative reasons (and correspondingly that moral claims are or entail claims about irreducibly normative reasons) (Olson 2014:124).

In order to fully understand (O₁) and (O₂), we need to know what a *favoring relation* is and what makes such a relation *irreducibly normative*. Let us start by considering what favoring relations (or “reason relations”) are. In general: a fact, *F*, is a reason for an agent, *S*, to perform an action, ϕ , *if and only if* *F* counts in favor of *S*’s ϕ :ing (ibid:118; Alvarez 2017:§1). Given this, reason relations are favoring relations that can be characterized as three-place predicates of the form *R* (*F*, *S*, ϕ). Suppose, for illustration, that Brad wants to know more about wolves. If so, the fact that the Wikipedia article “Gray Wolf” contains information about wolves counts in favor of Brad’s reading this article,

¹ Olson distinguishes the argument from queerness (singular) from and the queerness arguments (plural). In his terminology, *the argument from queerness* is a two-step argument for the claim that “[i]t is in the end ‘less paradoxical’ to reject than to retain the common-sense belief in moral properties and facts (Olson 2014:85)” and that error theory is, for this reason, more plausible than moral realism. The first step of the argument from queerness is to establish a presumption against moral facts (and properties). This is pursued by offering *queerness arguments* – i.e. arguments for the claim that moral facts would be queer and therefore ontologically suspicious. According to Olson, there are four queerness arguments to be found in Mackie’s writings, and thus four distinct versions of the argument from queerness. The second step of the argument from queerness is to explain why we think and speak as if there are moral facts, although there are none. This is pursued by defending projectivism or formulating debunking explanations of moral belief (Olson 2014:3, 84-6, 2016a:400, 2016b:463-4). For another evolutionary explanation of morality, see Joyce (2006).

and is therefore a reason for him to do so. Moral reasons follow this general schema; a fact is a *moral* reason for *S* to φ if and only if *F* is a *moral* consideration that counts in favor of *S*'s φ :ing.²

Premise (O₁) says that moral facts entail that there are facts that favor certain courses of behavior, *where the favoring relation is irreducibly normative*. So according to Olson, it is the favoring relation, *R*, persisting between *F* and *S*'s φ :ing that is irreducibly normative. The fact that is the reason (i.e., *F*) need not be irreducibly normative (Olson 2014:118). Moreover, (O₁) says that moral facts entail that there are facts standing in such relations – i.e., that there are irreducibly normative reasons. More specifically,

[...] moral facts are facts about what other facts (for example, the fact that performing some action would be conducive to the general happiness) favour certain courses of behaviour (for example, performing the action that would be conducive to the general happiness), where the favouring relation is irreducibly normative (ibid:135).

One of Olson examples of how moral facts entail irreducibly normative reasons runs as follows: Suppose that it is morally wrong to eat meat. The fact that it is morally wrong to eat meat entails that there is a reason not to eat meat. The reason – the fact that counts in favor of not eating meat – might be that eating meat is detrimental to human and non-human well-being (ibid:118-9).

Premise (O₂) says that irreducibly normative favoring relations are queer. But if a moral fact is a fact about another fact's standing in such a relation, it would seem that moral facts are also queer. Olson says that:

It seems difficult to deny that if the irreducibly normative favouring relation, or instances of it, is queer, then the *fact that it obtains* is also queer (ibid:135, my italics).

Given this, we may express the claim made in (O₂) by saying that *moral facts are queer* (because they entail that there are facts that are irreducibly normative reasons), or that *moral reasons are queer* (because they have the irreducibly normative property of counting in favor of actions) (ibid:117-8).³

² The term "reason" can be used in to two ways: (1) as a three-place predicate to refer to a *favoring relation* persisting between a fact, *F*, and an action, φ , of an agent, *S*. (2) as a one-place predicate to refer to *the fact that favors* (i.e., *F*) which constitutes a *relatum* of a favoring relation (the other *relata* being *S* and φ). I use "reason" in the sense of (2). I use "favoring relation" or "reason relation" to refer specifically to (1) (see also Bedke 2010:48 and Olson 2014:118).

³ This is to be understood as implying that if there are *non-moral facts* which *also* entail that there are facts that are irreducibly normative reasons, such facts are also queer. Olson says that "it is not morality in particular, but irreducible normativity in general, that is queer (2016b:462)." This will be important for the discussion in chapters 5 and 6.

Let us now consider what makes a favoring relation *irreducibly normative*. To understand this, we need to introduce a new distinction between different kinds of reasons. I said in chapter 3 that hypothetical (or “desire-dependent”) reasons are reasons for agents to take the means to satisfy their desires (etc.), and that desire-transcendent reasons are reasons for action that agents have regardless of whether they care to and regardless of whether performing the action satisfies any of their desires (etc.). Olson’s central distinction is rather between *irreducibly normative* and *reducibly normative* reasons. Desire-dependent reasons can be understood as belonging to the latter.⁴

A *reducible* reason relation is a favoring relation that reduces to empirical or otherwise non-normative facts – e.g., facts about what promotes desire satisfaction, about agents’ roles and rule-governed activities, or about correctness norms (ibid:121).⁵ Claims about such reasons can be fully analyzed in non-normative terms. Soccer players, says Olson, have reasons not to play the ball to their own goalkeeper while under pressure. This is a reducible reason because the claim that soccer players have such reasons might simply mean that performing this action is likely to provide the opponent team with opportunities to score. Similarly, soldiers have reasons to comply with the orders of their generals. These reasons are reducible because the claim that soldiers have reasons to comply with the orders of a general might simply mean that doing so is to comply with the orders of someone superior in military rank, which is part of the role of being a soldier. Because agents can occupy roles they have no desire to fulfill, or engage in activities they have no desire to succeed in, it is not the case that all reducible reasons are desire-transcendent. A soccer player who does not desire to win the game might still have a (reducible) reason not to play the ball to his own goalkeeper while under pressure. Similarly, a soldier may have (reducible) reasons to comply with the orders of a general *even if* he has no desire that would be served by his doing so (ibid:120-1).

An *irreducibly normative* reason relation, on the other hand, is a favoring relation that does *not* reduce to empirical or otherwise non-normative facts. It cannot be reduced, for example to facts about what would promote satisfaction of some of the agent’s desires, about what constitutes fulfillment of some role inhibited by the agent, or about what constitutes compliance with some rule (ibid:122, 135). Claims about such reasons *cannot* be fully analyzed in non-normative terms. More specifically:

⁴ I am not suggesting that Olson holds that all hypothetical reasons claims reduce to empirical claims. This is something Olson rejects (2014:152-5, 2016b:471). His view is rather that “hypothetical reasons claims are sometimes most plausibly interpreted as empirical claims about mean-end relations (2016a:401).”

⁵ More specifically: a fact, *F*, *reducibly* favors *S*’s ϕ :ing if (for example) *F*’s favoring *S*’s ϕ :ing reduces to the fact that ϕ :ing is conducive to the satisfaction of one of *S*’s desires, *D*, and ϕ :ing’s conduciveness to the satisfaction of *D* depends on *F*’s being the case.

a fact, F , is an irreducibly normative reason for an agent, S , to behave in a certain way, φ if and only if F counts in favor of S 's φ :ing where the favoring relation is irreducibly normative (see *ibid*:122).

We have seen that premise (O₂) of the argument from irreducibility says that that irreducibly normative favoring relations are *queer*. But what does it mean to say that something is queer? What does queerness more specifically consist in? (see Evers 2016:406) Olson's response to this is that "whatever is irreducibly normative is queer in virtue of being irreducibly normative (Olson 2016b:463)." Olson's worry with respect to irreducibly normative reasons is a *metaphysical* worry concerning (metaphysical or ontological) *irreducibility*. Error theorists can recognize reducible reasons – even those that do not depend on agents' desires. There is nothing metaphysically queer about conventional norms, rules, or standards of correctness that require or recommend various courses of behavior (Olson 2014:121). However, they cannot recognize irreducible reasons:

Irreducibly normative reasons are very different. [...] When the irreducibly normative favouring relation obtains between some fact and some course of behaviour, that fact is an irreducibly normative reason to take this course of behaviour. Such irreducibly normative favouring relations appear metaphysically mysterious. How can there be such relations? (*ibid*:136)

I agree with Olson when says that although he can imagine irreducibly normative facts, he finds it puzzling how there can be such facts (2016b:463). However, we saw above that Olson takes the queerness of irreducibly normative facts to establish a presumption against their existence. But why should queerness be a ground for such a presumption? It has been pointed out that there are many things that seem queer in one way or another, but whose queerness *does not* make us question whether they exist. Aardvarks, neutrinos and impressionist paintings, for example. Olson's response to this is that although aardvarks, neutrinos and impressionist paintings are *prima facie* queer, "they are [...] parts of the best explanations of some of our observations and beliefs (Olson 2014:87)." When we realize that this is so, we realize that these things are not queer in such a way as to make us doubt their existence (see *ibid*:86-8).

In this section, we have seen that Olson's conceptual claim is that moral facts entail that there are irreducibly normative reasons – i.e., that there are facts that favor certain courses of behavior, where the favoring relation is irreducibly normative – and correspondingly that moral *claims* entail *claims about irreducibly normative reasons*. We have also seen that Olson's worry

with respect to irreducible normativity is a metaphysical worry that specifically concerns favoring relations that do not reduce to non-normative facts, and reasons claims that cannot be analyzed in non-normative terms.

4.2 A challenge to the conceptual claim

It seems to me that there are two main strategies for critics of error theory to respond to the argument from irreducibility. One is to challenge the queerness-claim. A non-naturalist, for example, might argue that it is just a fundamental fact about reality that there are irreducibly normative favoring relations, and refuse to admit that such relations are queer.⁶ Another is to agree with Olson that we should not believe that agents can have irreducibly normative reasons for action, while insisting that moral discourse is not committed to the claim that there are such reasons, or at least not committed to it in a way that makes it involve a fundamental and serious error (see Olson 2014:125). A relativist, for example, might suggest that moral discourse should rather be analyzed along relativist lines. This type of response will be discussed below. Alternatively, it could perhaps be argued, along the lines of what was suggested in chapter 3, that if moral discourse involves a commitment to irreducibly normative favoring relations, the relevant assumption can be understood as entering pragmatically into moral conversations in such a way that its falsity does affect the truth values of moral judgments and statements. A third option, mentioned by Olson, is to argue that whether there are moral reasons for an agent to act in one way or other depends upon facts about her desires and interests. According to Olson, this is a deeply counterintuitive move:

[...] [O]ne cannot escape a moral reason to donate to Oxfam simply by adverting to one's lack of a relevant desire. Moreover, even if most or all people do have desires whose satisfaction would be promoted by their donating to Oxfam, or more generally by their compliance with moral norms, it is difficult to accept [...] [the] view that whether there are reasons for people to donate to Oxfam, or more generally to comply with moral norms, depends on whether doing so would promote fulfilment of their desires and interests. This would simply be the wrong explanation of why there are reasons to donate to Oxfam, or more generally to comply with moral norms (ibid:125).

Olson's point here seems to be that given that there are moral reasons for an agent to donate to Oxfam, the *right* kind of explanation of why this is so is that donating to Oxfam is the morally right thing to do (see ibid:118). As we have seen, Olson takes moral facts, such as the (possible) moral fact that donating to

⁶ See e.g. Toppinen (2016:443 ff.). For discussion, see Olson (2014:136 ff., 2016b:463 ff.).

Oxfam is the morally right thing to do, to entail that there are facts that are irreducibly normative reasons for agents to act in some ways rather than others. And it is in virtue of their having this feature that moral facts are queer.

Olson's main argument in favor of the claim that moral claims entail claims about irreducibly normative reasons is a response to Finlay's (2006, 2008) relativistic account of moral thought and discourse.⁷ Olson challenges Finlay's (2008) argument against absolutism (i.e., non-relativist accounts of moral discourse), and argues that evolutionary considerations favor the conceptual claim over Finlay's suggestion that moral claims are rhetorical devices used to put pressure on people's actions.

Olson's focus on Finlay's proposal might be motivated by the idea that to the extent that reductive accounts of normativity encounter problems, there is reason to think that we are committed to irreducible normativity, and the fact that Finlay's proposal is one of the more powerful recent reductive accounts of normativity. If Olson is right that this proposal fails, he would probably think that the burden of proof lies with the opponent. But it can be argued that this misrepresents the evidential situation. Our failure to have any *reductive* normative content in mind *does not* imply that we have some irreducible content in mind, nor that we take normativity to be irreducible. If all reductive analyses fail, this might be because we have no clear content in mind, with various strands of our thinking pulling in different directions.⁸

In the following sub-sections, I challenge Olson's semantic claim that moral claims entail claims about irreducibly normative reasons by challenging parts of his response to Finlay. Even granting that irreducibly normative favoring relations are (or would be) queer, we have reasons to believe that moral claims entail claims about queer relations only if we have reasons to believe that moral claims entail claims about irreducibly normative favoring relations. If I am right, there is some reason to doubt that Olson (2014, 2016a, 2016b) succeeds in providing such reasons.

Finlay's relativism

Finlay defends relativism (or "contextualism") with respect both to the metaphysics of moral values and to the semantics of moral claims. According to him, "every kind of value is relative to some standard or end (Finlay 2008:350)," and moral claims are always relativized to a (contextually implicit) end or system of ends. An end, *E*, is "a possible aim for action or object of desire (Finlay 2006:8)." The suggestion that moral claims are always relativized to

⁷ At least, this is how I understand matters. Olson says that "The most elaborate attack on the conceptual claim is developed in Stephen Finlay's recent work. Seeing where it goes wrong will serve to illustrate the costs associated with rejecting the conceptual claim [...]. This is the task in the following section (Olson 2014:125-6)." The section he refers to in this passage is called "In defense of the conceptual claim".

⁸ Thanks to Gunnar Björnsson for pointing this out.

ends is not the suggestion that we always use sentences of the form “in order that *E*, you (morally) ought to φ ” to make moral statements, but rather that this is how moral statements should be *analyzed*. According to Finlay (2008:353), the reason why we use sentences that leave out the relativization – e.g. sentences of the form “you (morally) ought to φ ” – to make moral claims, is that we rely on the context to provide the relevant end. The reason why we can rely on context to do so is that moral ends or standards are easy to identify, and are typically shared by members of a society.

Finlay (2006:7-9) offers a naturalistic reduction of favoring relations according to which reasons are “end-relational” explanatory relations: a fact, *F*, is a reason for φ :ing, relative to a system of ends *E*, *if and only if* *F* explains why φ :ing is conducive to *E*. The fact that is the reason (i.e., *F*) is a fact that explains why the action in question is conducive to the end in question. According to this view, a reasons claim is a claim about an end, an action, and a fact (where the fact is such that it explains why the action would be conducive to the end). Consequently, reasons claims can be analyzed in terms that are not themselves normative.

Moreover, Finlay suggests that whether or not a reason *matters* to an agent depends on her cares and concerns.⁹ His point seems to be that while there might be a desire-transcendent moral reason (e.g.) for Mark to tell his girlfriend that he cheated on her last weekend, this matters to Mark *only if* he embraces the (contextually implicit) moral end to which telling his girlfriend the truth is conducive.¹⁰ (The end might, in this case, be that Mark treats his girlfriend with the respect she deserves.) If Mark does not embrace the relevant moral end, or does not accept moral ends at all, this reason might not matter to him (Finlay 2006:17-18, see also Olson 2014:126).

Some of the implications of this relativistic view of normative authority resemble the implications of my suggestion, in chapter 3, that moral practice can be understood as an institution. Finlay says that his view implies that

[m]oral reasons matter (intrinsically) only for those with moral concerns [...]. A morally unconcerned person can still have moral reasons, but they cannot be reasons that matter for him (Finlay 2006:17-8).

I claimed, above, that the suggestion that moral practice is an institution entails that strictly speaking, moral norms have (practical) authority only for agents who endorse the moral institution. I also claimed that because most of us do

⁹ Finlay expresses this point as follows: “[...] all genuine value has only contingent authority, which exists only for those with the right psychological orientation towards the relevant standard or end (Finlay 2008:351).”

¹⁰ The fact that is the reason might still matter to Mark in relation to *other* ends. I will ignore this complication here.

endorse the moral institution (exceptions are very rare), most of us do take morality seriously, and do take moral norms as guide to our behavior. Finlay might have something similar in mind when he says that

[w]hat is jarring about Williams's view is the suggestion that the monster who is untouched by moral concern isn't ignoring any reasons that matter [...]. [...] Williams errs here: the monster violates moral reasons, and because *we* are motivated by moral concerns, these reasons (and this violation) matter *for us*, and therefore it is appropriate for us to judge that he violates 'important' reasons, and generally inappropriate to deny it.

I contend that this is sufficient to accommodate our condemnation of the monster [...]. 'The normative force of moral judgement', one may object, 'requires that an agent's immorality violates reasons that matter for that very agent'. But this strikes me as merely an expression of the futile hope that there be some philosophical means of persuading every individual to virtue, or minimally, that every villain has failed to recognize important truths (Finlay 2006:18).

It transpires from this (intentionally sparse) account of Finlay's relativistic theory that if Finlay is right, Olson attributes an error to moral discourse that simply is not there. According to Finlay's account, the reasons claims entailed by ordinary moral claims can be fully analyzed in non-normative terms; i.e. ordinary moral claims *do not* entail claims about irreducibly normative favoring relations (see Olson 2014:126).

Finlay's challenge to absolutism

Olson's response to Finlay is a response to what Olson (2014:127) interprets as a challenge to the conceptual claim.¹¹ Finlay's challenge is directed at "the disputation evidence" in favor of absolutist (non-relativistic) interpretations of moral discourse. The evidence consists in the observation that we sometimes engage in moral discussions with people whose fundamental moral views differ (or seem to differ) from ours, and that when we engage in such discussions, we do so with the aim of convincing our opponents that we are right and they are wrong. This observation counts against relativism because if moral claims are relativized to ends (and speakers recognize this to be the case), it is hard to explain why we engage in such moral discussions with such aims: "[w]ere

¹¹ Finlay's (2008) argument is directed at Mackie/Joyce argument for moral error theory. He interprets this argument as resting on the assumption that moral discourse presupposes that moral value has absolute authority (over persons). Finlay argues that his account of moral reasons entails that moral value has merely *relative authority* since whether a moral reason *matters* to the agent depends upon the agent's cares and concerns.

people to recognize that they subscribe to fundamentally different moral standards, then they should recognize that there is no genuine dispute between them (Finlay 2008:355).”

Finlay (2006:13, 2008:357) argues that relativism can explain these observations. He suggests that when we engage in moral discussions with people who do not share our fundamental moral values, we withhold the relativization of our moral claims as a pragmatic device to win our opponents over. According to him, behaving as if an end is presupposed in a context where it isn't rhetorically expresses the demand that one's audience share or respect that end. By withholding the moral standard or end to which her claim is relativized, a speaker rhetorically expresses an expectation (demand) that her opponent subscribes to her standard or end. If this is right, the disputation evidence does not undermine relativistic analyses of moral discourse.

Before examining Olson's response, it is worthwhile considering the current dialectic. While it is clear that Finlay's and Olson's analyses of moral discourse cannot both be correct, it is not clear to me that a successful rebuttal of Finlay's attack on the disputation evidence can provide Olson's conceptual claim with any strong support.

The claim Finlay attacks by attacking the disputation evidence is not specifically Olson's claim that moral claims entail claims about irreducibly normative reasons (or that moral *facts* entail irreducibly normative reasons). The “disputation evidence” is evidence in favor of absolutist analyses of moral claims, so by attacking it, Finlay attacks *any* analysis of moral claims according to which such claims are not analyzed as relative to ends or some other parameter.¹² Olson's thesis is a version of absolutism, so Finlay's attack certainly extends to it. However, if Finlay's attack is directed at absolutism rather than specifically at Olson's conceptual claim, it seems to me that while a successful rebuttal of Finlay's attack strengthens absolutism over relativism, it does not give Olson's conceptual claim any support *over and above* the support *any* absolutist thesis gains from such a rebuttal. Since there are many thinkable varieties of absolutism – some of which do not involve the claim that moral claims entail claims about irreducibly normative reasons – strengthening absolutism does not give us strong reasons to believe that Olson's absolutist theory is true.¹³ (I am not hereby suggesting that Olson takes this to be the case.)

¹² Finlay discusses the disputation evidence in a section titled “The Evidence for Absolutism”, where he argues that “[...] a nonabsolutist, relational interpretation of moral discourse can explain the evidence at least as well, [...] therefore [...] we should not embrace the absolutist, nonrelational interpretation [...]” (Finlay 2008:352).”

¹³ Matt Lutz and Stephen Finlay raise a similar worry with respect to Olson's argument for the claim that Finlay's relativist theory of moral thought and discourse is unsuccessful: “Finlay's theory is only one of many versions of naturalism [...] that deny a commitment to ‘irreducibly normative’ reasons, so its refutation would lend the premise scant support (Lutz and Finlay 2016:1224).”

Olson's response to Finlay

Let us now consider Olson's response to Finlay's attack on the disputation evidence. According to Olson, Finlay's explanation of why we engage in moral discourse with people who do not share our fundamental moral values backfires. He says that

[t]he idea that moral judgements are partly rhetorical devices used to put pressure on people to behave in certain ways is congenial both to moral error theory and to Finlay's relativist theory, but it *fits better with the former* (Olson 2014:128, my italics).

Olson supplies two arguments in favor of the claim that Finlay's suggestion that moral claims are partly rhetorical devices used to put pressure on people's behavior fits better with error theory than with relativism. The first is that Finlay's suggestion

[...] fits well with a hypothesis congenial to moral error theory, namely that part of the reason why moral thought and talk evolved is their coordinating and regulative functions that are highly useful from an evolutionary perspective. [...] It is a plausible conjecture that moral discourse fulfils these functions better if moral claims entail claims about irreducibly normative reasons, than if they are reduced to claims about what would conduce to some end (Olson 2014:128-9).

As an initial comment, it seems to me that the hypothesis that the evolution of moral thought and talk can partially be explained with reference to its regulatory and coordinating functions is congenial not only to Olson's thesis that this function is carried out by the semantics of moral claims, but also to Finlay's thesis that it is carried out by the pragmatics of moral claims. To me, the claim that a certain type of discourse evolved because it serves a function useful to us does not say anything about *how* such discourse serves this function. Performatives, for example, supposedly also evolved because they serve a function useful to us, but I fail to see how knowing this could help us understand *how* performatives perform their function.

More important issues concern Olson's suggestion that moral discourse performs its regulatory and coordinating functions better if moral claims entail claims about irreducibly normative reasons. My first issue with this suggestion is that I do not see what reasons there could be for thinking that it is true. Unfortunately, Olson does not elaborate on this matter. Is it, for example, a general truth that a statement's entailments have a stronger power to influence the behavior of the agent addressed than any rhetoric or otherwise non-

semantic aspects of the statement? It seems to me that as far as a statement's power to create its desired effect is concerned – be it an action, an emotion, or a belief – the vital issue is what message is being communicated, regardless of *how* this message is communicated.

My second issue with this claim is that even granted that moral discourse performs its regulatory and coordinating function *better* if moral claims entail claims about irreducibly normative reasons, we can explain how it is that moral discourse has this power even without supposing that moral claims entail claims about irreducible normative reasons. Whether or not one agrees with Finlay that moral values are typically shared, and that we are for this reason able to express demands by leaving out the end to which our moral claims are relativized, one might believe that normally when we make a moral statement, we operate under the assumption that all parties to the conversation give practical weight to moral requirements and reasons, and give moral considerations precedence in practical intra- and interpersonal deliberation. By behaving in this way, we signal that we take seriously the moral considerations brought to attention by our statements, thereby putting pressure on our audience to do the same.

I also believe that it can be argued that our addressees normally put pressure on themselves to take seriously any moral claims we direct at them. I am ready to admit that *if* people in general did not care about moral values, acting in a morally commendable fashion, treating others right, etc., then a widespread belief in Finlay's claim that moral reasons matter only to those who care about the ends of morality would probably diminish the regulatory and coordinating powers of moral talk. But people are not like that. As I suggested in chapter 3, above, we generally *do* care about morality, and generally find it important that people (ourselves and others) take moral norms as guide to their behavior. For this reason, when others ascribe moral reasons to us, inform us that there is some moral requirement that applies to us in our current situation, remind us of our moral duties, etc., we take these claims seriously. Moreover, because we care about morality, we refuse to admit that someone like Gyges can escape the requirement not to kill innocent people simply by renouncing the moral community. If we come across someone like Gyges, we will insist that he has a reason to respect other people's rights to their own lives, bodies, possessions, and so on, and continue to put pressure on him to act in a morally upstanding fashion. Finlay could perhaps put the point as follows: because we embrace the ends of morality, moral claims matter to us, and this is sufficient for us to take moral discourse seriously enough for such discourse to serve its regulatory and coordinating function.

Olson's second argument in favor of the claim that Finlay's suggestion that moral claims are partly rhetorical devices fits better with error theory than with relativism is that the most straightforward explanation why moral claims have this rhetorical force is that moral claims entail claims about irreducibly

normative reasons: “moral claims have rhetorical force *because* they are or entail claims about irreducibly normative reasons (Olson 2014:129).”

It is a common ground between Finlay and Olson that moral claims have more rhetorical force than (e.g.) etiquette claims, and that we take moral appraisals more seriously than appraisals of etiquette. Their disagreement concerns how this observation is to be explained. Finlay’s explanation is reminiscent of a point I made above, namely that we care more about morality than about other normative systems (like etiquette). He says that

[m]oral standards or ends are of pressing concern to [...] [us], and their importance to us typically overrides the importance of other standards and ends. This explain[s] why we are much more serious and intransigent about our moral appraisals (Finlay 2008:354).

Olson’s objection to this is that

[...] one would expect the difference in seriousness and intransigence between moral claims and etiquette claims to be reflected in the concepts we use to make them. The conceptual claim makes good on this expectation: the fact that moral standards or ends are of especially pressing concern to us explains why moral claims entail claims about irreducibly normative reasons (Olson 2014:129).

This reply strikes me as question-begging. Given that the question is how the observation that there is a difference in seriousness and intransigence between moral claims and etiquette is to be explained, and that Olson’s favored answer is that it is to be explained with reference to the concepts used in making moral claims, the claim that “one would expect the difference [...] to be reflected in the concepts we use to make them (ibid)” does not sound like a good argument for the claim that the difference is best explained with reference to the concepts used in making moral claims. (Perhaps I am missing something essential here. Perhaps, there is some implicit premise or argument that explains why we should agree with Olson that one would expect our difference in attitudes to be reflected in the concepts used. Is the analogue of this claim true with respect to other associated phenomena where there is a general tendency to take the one more seriously than the other?) Moreover, even given that the difference in seriousness and intransigence between moral claims and etiquette claims is reflected in the concepts involved in these claims, it is not clear to me why irreducibility would go with more seriousness, and reducibility with less. Why, that is, would the referents of the concepts involved in those judgments and

claims we take more seriously and are less willing to give up be irreducible? What does irreducibility have to do with seriousness?

In addition to the replies to Finlay's treatment of the disputation evidence discussed above, Olson formulates an argument for the claim that Finlay's relativist theory of moral thought and discourse is unsuccessful.¹⁴ For all I have suggested above, Olson may be right that relativism fails. Because I do not attempt to defend relativism here, and because I do not consider this argument as obviously relevant to determining the merits of Olson's conceptual claim that moral claims entail claims about irreducibly normative reasons, I will leave this argument as it stands.

4.3 Conclusion

This chapter has discussed what I have referred to as *the argument from irreducibility*, which has been suggested by Jonas Olson. I challenged the argument by challenging Olson's argument for the claim that moral claims entail claims about irreducibly normative reasons. Even granting that irreducibly normative favoring relations are, or would be, queer, moral claims entail claims about queer relations only if they entail claims about irreducibly normative favoring relations.

If my arguments are sound, there is reason to doubt that moral claims entail claims about irreducibly normative reasons. If so, the argument from irreducibility has not convincingly shown that moral thought and discourse involve an error of the kind that makes all positive moral statements and judgments untrue.

¹⁴ According to Olson, Finlay's view encounters several related problems having to do with *fundamental moral claims*, and claims to the effect that some moral standard is the correct fundamental moral standard (see Olson 2014:130-5, 2016b:465-467). For discussion see e.g. Evers (2016:410-415).

Chapter 5

The epistemic objection

In this and the following chapter, I discuss two objections to ERROR THEORY. Both objections concern what implications (if any) ERROR THEORY has within the field of epistemology and for epistemologically relevant states and relations, i.e. for things such as justified belief, epistemic reasons, evidence, for what it is for someone to believe or rationally believe some proposition, what it is for something to be a good reason or good evidence for some belief, etc. The first objection, which I refer to as the *epistemic objection*, is suggested by Terence Cuneo (2007). The second, which I refer to as the *psychological objection*, draws upon an argument suggested by Bart Streumer (2013, 2016a, 2016b). (Streumer himself does not defend the psychological objection, and does not intend to raise any other objection to error theory, but defends one of its premises for purposes of saving normative error theory from objections.) I argue that while both objections have their merits, neither objection gives us strong reasons to believe that ERROR THEORY is incorrect. In particular, none of them convincingly shows that ERROR THEORY has implications that the error theorist cannot handle.

The topic of this chapter is the epistemic objection. The psychological objection is the topic of the next chapter. According to the epistemic objection, moral error theory entails an epistemic analogue to it – *epistemic error theory*. Epistemic error theory involves the claim that there are no epistemic reasons, which is claimed to have several problematic consequences – including that there can be no good arguments for error theory, that error theory cannot be rationally believed, and that error theory is polemically toothless in metaepistemological and metaethical debates (Cuneo 2007:117-122). Consequently, if the epistemic objection is successful, moral error theory has problematic consequences. The entailment claim can be motivated as follows: Moral error theorists argue that (positive) moral statements are true only if moral reasons have property *f*, and that moral reasons do not (or cannot) have property *f*. But (positive) *epistemic* statements can be true only if *epistemic* reasons have *f*. Consequently, if the arguments for moral error theory are successful, epistemic error theory is true. The epistemic objection can be understood as “a companions in guilt”-response to arguments for moral error

theory. If the arguments for moral error theory (if successful) show that *epistemic* error theory is true, and epistemic error theory is false (or we have good reasons to believe so), the arguments for moral error theory prove too much (see Fletcher 2009:363-4; Cowie 2014:407-8; Olson 2014:11).¹ Given this, the epistemic objection, if successful, does not merely show that moral error theory has problematic consequences, but also that the arguments for it are no good.

Below, I raise two challenges to the epistemic objection. First, I indicate some possible responses to the challenges to epistemic error theory. If these responses are successful, there are ways for moral error theorists who accept the entailment claim to avoid (some) of the problematic consequences of epistemic error theory. Second I give an account of epistemic reasons according to which such reasons are hypothetical and hence relevantly *different* from moral reasons. Given this, the arguments for moral error theory can be successful without establishing the truth also of epistemic error theory. So the entailment claim is false.

I proceed as follows: First, I give an account of how the entailment claim can be defended (section 5.1). Second, I describe some challenges to epistemic error theory, and indicate some possible responses (section 5.2). Third, I challenge the entailment claim by giving an account of Hilary Kornblith's suggestion that epistemic reasons are hypothetical reasons, and by defending it against some objections (section 5.3).

5.1 The entailment claim

The argument for the claim that moral error theory entails epistemic error theory can schematically be represented as follows (from now on, “the entailment claim” refers to the claim that moral error theory entails epistemic error theory, unless otherwise indicated):

- (1) [p] Positive moral statements are true only if things have feature *f*,
- (2) [p] Things do not (or cannot) have *f*,
- (3) [1, 2] There are no true positive moral statements,
- (4) [p] Positive epistemic statements are true only if things have *f*,
- (5) [4, 2] There are no true positive epistemic statements.

¹ For more on companions in guilt-arguments, see Lillehammer (2017). Cowie (2014) argues that companions in guilt-arguments against moral error theory that take epistemic facts as companions in guilt for moral facts suffer from an internal tension.

The (1)-(5) argument targets *specifically* those moral error theorists who defend moral error theory with an argument of the (1)-(3) kind.² The idea is that if such an argument shows that moral error theory is true, it will also show that epistemic error theory is true. Hence, the entailment claim is true. The nature of the *things* and the *feature, f*, mentioned in (1) and (2) is determined by the particulars of the specific argument for moral error theory to which the error theorist targeted by the (1)-(5) argument subscribes. The things might be moral reasons, moral requirements, moral facts, etc., and *f* might be non-institutionality, irreducible normativity, objective validity, etc.³ The arguments from non-institutionality and irreducibility, discussed in chapters 3 and 4, can both be understood as (1)-(3) arguments for moral error theory, each focusing on a particular feature of moral reasons – viz. *non-institutionality* and *normative irreducibility*. For this reason, the (1)-(5) argument for the entailment claim can be modeled on either of them; i.e., can be directed at error theorists who subscribe to either of these arguments.

If, for example, the (1)-(5) argument is modeled on the argument from irreducibility, its details can be filled in as follows: According to the best argument for moral error theory – viz. the argument from irreducibility – positive moral statements are true only if there are moral reasons, and there are no moral reasons. Moral and epistemic reasons are analogous in the sense that that feature of moral reasons on the basis of which error theorists have concluded that moral reasons do not exist – viz. *normative irreducibility* – is also a feature of epistemic reasons, if they exist. Positive epistemic statements are true only if there are epistemic reasons, so no epistemic statement is true.⁴

I suspect that on most versions of the (1)-(5) argument, the *things* in question will be moral reasons, and *f* some feature of moral reasons. This suspicion is motivated by the following two observations: First, two of the most well-known and currently most debated arguments for moral error theory are arguments according to which moral discourse presupposes that moral reasons have a particular feature *f*, and *f* is hard to make sense of, irreconcilable with the way things are, etc. Second, it has been argued that moral and epistemic reasons are structurally similar. We saw in chapter 4 that it has been suggested that

² It is possible that someone accepts moral error theory on the basis of some other kind of argument, e.g. an argument for the claim that all alternative metaethical views fail (see Joyce 2011a:520, 2016a:§4). Consequently, the (1)-(5) argument, if successful, does not show that *all* moral error theorists are committed to epistemic error theory.

³ Depending on how the moral error theorist fills in the details of (1), and on how she motivates (2), offering a (1)-(3) argument may be compatible with holding that *things other than those which constitute the target of the (1)-(3) argument have f*. On one reading of Joyce, he defends moral error theory on the basis of a (1)-(3) argument where the “things” are *moral reasons* and *f* is the feature of *strong categoricity* or of *being real reasons for action*. Joyce investigates (and rejects) the possibility that the reasons delivered by practical rationality have this feature, thus presupposing that it is possible that although the relevant feature does not belong to the things serving as target of his argument, it does belong to other things (Joyce 2001:51 ff.).

⁴ This kind of argument for the entailment claim is discussed by Olson (2011a:80-2, 2014:155-6) and Rowland (2013:3 ff.). See also Bedke (2010:42 ff.).

moral reasons are (or would be) moral facts or considerations that *count in favor* of certain actions or behavior. Given this, and that moral and epistemic reasons are structurally similar, epistemic reasons are (or would be) epistemic facts or considerations that count in favor of forming certain beliefs, and of other doxastic behavior such as revising or abandoning beliefs (Olson 2014:155). If the (1)-(5) argument is modeled on the argument from irreducibility, both moral and epistemic reasons are moreover such that the favoring relation between a fact and an agent's actions or behavior is irreducibly normative. Consequently, so this formulation of the argument for the entailment claim goes, if the argument from irreducibility successfully casts doubt on relations of irreducible favoring in the moral domain, it unavoidably also casts doubt on such relations in the epistemic domain (ibid).

5.2 Cuneo's challenges to epistemic error theory

In this section, I accept the claim that moral error theory entails epistemic error theory for the sake of argument. (The purpose of the next section, 5.3, is to raise a challenge to this claim.) The purpose of the present section is to discuss some challenges to epistemic error theory and indicate some possible responses.

Cuneo offers an extensive argument against epistemic error theory.⁵ According to his account, epistemic error theory is an epistemic analogue to error theory, and involves claims analogous to COGNITIVISM and FAILURE (Cuneo 2007:8, 21-8, 115-6): it involves the claim that epistemic judgments are beliefs (with epistemic propositions as their content), that epistemic utterances are assertions of epistemic propositions, and that no (positive) epistemic proposition is true. So why might one think that epistemic error theory follows from moral error theory? According to Cuneo, epistemic reasons are analogous to moral reasons (Cuneo 2007:38, 59, 62), and epistemic error theory involves the claim that there are no epistemic facts – i.e., an epistemic analogue of ANTI-REALISM (ibid:116).⁶

Cuneo's objections to epistemic error theory (from now on, "error theory" and "error theorist" refers to *epistemic* error theory and epistemic error theorists, unless otherwise indicated) starts from the claim that because the

⁵ It might seem that Cuneo can be interpreted either as providing arguments to the effect that epistemic error theory is false, or as providing arguments merely to the effect that someone who adopts or accepts epistemic error theory finds herself in a theoretically and dialectically costly position. This impression, I believe, is mistaken. Cuneo's goal is to defend moral realism. One of the premises in Cuneo's argument for moral realism is the claim that epistemic facts exist. Cuneo defends this premise by attacking epistemic error theory and other theories according to which there are no epistemic facts (Cuneo 2007:6-8). Given this, Cuneo's challenges to epistemic error theory should be understood as challenges to its truth.

⁶ I suggested in chapter 2 that ANTI-REALISM is not a necessary component of ERROR THEORY. Given this, moral error theorists who do not accept ANTI-REALISM can ignore objections that presuppose that error theorists are committed to this claim or, perhaps by implication, its epistemic analogue.

error theorist must claim either that *there are* epistemic reasons to believe error theory, or that *there are no* such reasons, she faces the following dilemma:

Either epistemic nihilism is self-defeating and hence, we have no (sufficient) reason to believe it, or, it implies that there are no epistemic reasons and, a fortiori, that we have no reason to believe it (ibid:118).⁷

According to Cuneo (ibid:117), holding that *there are* epistemic reasons to believe error theory is not a viable option for the error theorist. The reason for this is that it is that error theory involves the claim that there are no epistemic facts, which is incompatible with the claim that there are epistemic reasons; accepting both claims would render the error theorist's belief system incoherent. It is reasonable to suppose that facts about epistemic reasons are epistemic facts. If there are epistemic reasons for an agent, *S*, to believe some proposition, *p*, there is *at least one* epistemic fact – viz. the fact that *S* has epistemic reasons to believe that *p*. If so, the claim that there are no epistemic facts implies that there are no facts about epistemic reasons. Consequently, if the error theorist claims that there are epistemic reasons to believe error theory, her position is self-defeating (Cuneo 2007:118).

Given this, it would seem that the obvious choice for the error theorist is to hold that *there are no* epistemic reasons to believe error theory. But according to Cuneo, this is not a viable option either. If the error theorist accepts this claim, she subjects herself to several related problems. I will discuss those three which seem the most serious.⁸

A first problem is that the error theorist commits herself to a “sweeping form of epistemological skepticism (Cuneo 2007:119).” Cuneo argues as follows:

[...] [the claim that there are no epistemic facts or reasons implies] that a sweeping form of epistemological skepticism is true. For if there are no epistemic facts or reasons, then none of our propositional attitudes can exhibit epistemic merits or

⁷ Cuneo uses “epistemic nihilism” to refer to epistemic error theory (Cuneo 2007:115), and “reasons to believe” to refer to epistemic reasons to believe (ibid:117).

⁸ In addition to the challenges discussed below, Cuneo mentions that the error theorist might seem to be committed to a Moorean paradox of the following sort:

(1) Error theory is true, but there is no reason to believe error theory.

However, he ultimately does not claim that the error theorist is committed to (1). The reason for this is that the error theorist accepts the claim that no belief can be epistemically justified or warranted. Consequently, according to her view, a person's accepting the second conjunct of (1) cannot undercut her epistemic justification for accepting the first. If error theory is true, (1) does not even rise to the level of paradox (Cuneo 2007:119, for discussion, see Olson 2014:165-6). Cuneo (ibid:121-2) also mentions that *if* belief attributions and meaning-attributions are normative in character, this creates further costs of epistemic error theory, but that his “strategy in this chapter has not been to defend these more controversial claims [...] [but] simply to indicate some of the more obvious and unattractive implications of epistemic [...] [error theory] (ibid:122).” The issue of whether it is possible to believe error theory will be discussed in the next chapter.

demerits; none of our propositional attitudes can be justified, warranted, entitled, irrational, a case of knowledge, based on reasons, or the like (ibid).

Commitment to epistemological skepticism is a cost because it is part of our commonsensical conception of the world that epistemological skepticism is false; its denial is something that mature adults take for granted in their everyday lives (ibid:123). The point is not that this observation is in itself sufficient to prove that epistemological skepticism is false, but rather that this creates presumption against any view that implies that it is true.

The epistemological skepticism Cuneo has in mind seems identical to the claim that no entity can display any epistemic merit or demerit. Given this, the error theorist might want to object that commitment to this thesis should not be treated as a *further* problematic implication of her theory, over and above commitment to the claim that there are no epistemic merits or demerits; these are merely two ways of phrasing one and the same implication. To this, Cuneo may respond that listing the variety of epistemic facts highlights the range of things that the error theorist has to give up, and thereby makes the costs of her position clearer.

The error theorist may attempt to rebut this challenge by “biting the bullet”. That is, she may admit that error theory entails that no propositional attitude can have an epistemic merit or demerit, but insist that this is a benefit of the theory. Since there are no epistemic facts, any theory that entails that *there are* epistemic merits or demerits has false entailments. This reply would, of course, have to be supplemented with a good argument for the claim that there are no epistemic facts. But it seems to me that it can be argued that whether this is a good reply to the present challenge also depends on whether the error theorist is able to provide a good explanation of *why*, given that no propositional attitudes display epistemic merits or demerits, *it seems to us that they do*. Why, for example, does it seem so obvious to us that some beliefs – e.g. someone’s belief that the Earth is flat or that the Ness Monster is real – are unjustified? Commitment to the claim that there are no epistemic merits or demerits is a cost of error theory because it seems to us that propositional attitudes *can* display such features. Hence, it seems that the error theorist who responds to the present objection by “biting the bullet” owes us an explanation of why appearances are misleading in this regard.⁹

A second problem facing the error theorist who holds that there are no epistemic facts or reasons is that this commits her to the claim that there can be no (epistemically) *good arguments* for anything. Cuneo says that

⁹ Rowland (2013:2, 13-14) similarly argues that if there are no epistemic reasons for belief, no one knows anything. According to him: If *S* knows that *p*, there is some epistemic justification for *S* to believe that *p*, and if there is some epistemic justification for *S* to believe that *p*, then there is an epistemic reason for *S* to believe that *p*.

[...] in the paradigmatic case, the premises of an argument are offered in support of its conclusion in the sense of providing evidential support for its conclusion. [...] A statement's being offered as evidential support for a conclusion, however, is just a matter of its being offered as a *reason* for accepting that conclusion. And, when all goes well, premises *are* reasons to accept a conclusion. [...]

[...] But [...] if nihilism were true, it would be impossible that there were premises of an argument that provide evidential support for its conclusion (Cuneo 2007:121).

Given this, the claim that there are no epistemic reasons implies that there are no good arguments. This has two important implications. First, the error theorist cannot provide any good arguments in favor of error theory without having to choose between accepting a contradiction – viz. that *there are* and *are not any* good arguments for error theory – and abandoning a claim central to error theory – viz. that there are no epistemic reasons. Neither appears to be an attractive option. Second, any argument in favor of the claim that error theory is false would beg the question against it (Cuneo 2007:119-121).

As Cuneo (ibid:121) also points out, the error theorist may attempt to rebut this challenge by suggesting an alternative understanding of the nature of good arguments. This line of response has been suggested by Olson (2014:157, 159), who proposes a distinction between arguments to the effect *that some proposition, p, is true*, and arguments to the effect *that there are reasons to believe that p*.¹⁰ While arguments of the latter kind presuppose or entail that there are epistemic reasons for belief, those of the former kind do not. With this distinction in place, the error theorist may attempt to reduce the force of the present challenge by arguing that although error theory entails that there cannot be any arguments that are epistemically good in the sense Cuneo describes, it does not entail that there cannot be any argument to the effect that propositions are true.

I agree with Olson (ibid:161-3) that to offer an argument for some claim can be an attempt merely to demonstrate that the claim in question is true. In such cases, the relation between premise and conclusion is best understood in terms of truth, or something closely related to it, such as truth-preservation. For example, it could be argued that to offer a statement, *s*, as premise in an argument for a conclusion, *c* is to offer *s* as being related to *c* in the following way: if *s* is true, *s* (taken together with any further true statements also offered as premises in the argument) guarantees the truth of *c*. When all goes well, *s* is true, and *does* stand in this relation to *c*. In such a case, the argument is logically sound, and *c* is true.

¹⁰ This distinction is also made by Guy Fletcher (2009:366).

Using Olson's distinction, the error theorist may put forward arguments to the effect that error theory is true without having to choose between accepting a contradiction and abandoning a claim central to her theory, and her opponent may offer arguments to the effect that error theory is false without begging the question against the theory (Olson 2014:157-9).

A third problem facing the error theorist who holds that there are no epistemic facts or reasons is that by accepting it, the error theorist renders herself polemically toothless in metaethical and metaepistemological debates. If a belief is epistemically *rational* to the extent that it conforms to (what the agent takes to be her) epistemic reasons, and *irrational* to the extent that it does not so conform,¹¹ the claim that there are no epistemic reasons entails that no belief – including the error theorist's belief that error theory is true, and her opponent's belief that error theory is not true – is rational or irrational. Given this, the toothlessness challenge can be phrased as follows:

If [...] epistemic nihilists hold that we do not have epistemic reasons to believe their position, then their position is polemically toothless in the following sense: No one would make a rational mistake in rejecting it and no one would be epistemically praiseworthy in accepting it (Cuneo 2007:117).¹²

Alternatively, the challenge can be phrased as the following dialectical problem: The error theorist does not believe that we have reasons to care about her arguments for error theory. She accepts this, of course, but *we* do not. We, who are not (yet) error theorists, or have not yet made up our minds as far as the truth of error theory is concerned, can notice that (i) if she is *wrong*, her arguments are no good, and (ii) if she is *right*, her arguments are no good either (because if error theory is true, there is no such thing as a good argument). For this reason, we can conclude, *prior to listening to the error theorist's arguments*, that her arguments are no good, and consequently that we are not missing out on arriving at better justified beliefs by nor listening to her.¹³

Olson suggests several ways in which the error theorist may attempt to rebut this challenge. I will discuss merely one of them here. Recall that Olson distinguishes between arguments to the effect that *p* is true and arguments to the effect that there are reasons to believe that *p*. In response to Cuneo's toothlessness challenge, he suggests that

[...] since error theorists are not in the business of offering arguments about what would be rational to believe or about

¹¹ This seems to be how Cuneo (2007:13, 58, 117) uses the terms "rational" and "irrational".

¹² Cuneo (2007:117) uses "rational" and "irrational" to refer specifically to *epistemic* rationality and irrationality.

¹³ Thanks to Gunnar Björnsson for pointing this out.

what there is epistemic reason to believe, there is no harm in conceding that error theory is toothless in those debates. What is important is that error-theoretical arguments have bite in debates on where the truth lies in metaethics and metaepistemology; these are the debates with which error theory is concerned. And given that the aim of metaethical and metaepistemological inquiry is to get at the truth, error theory is not polemically toothless in these debates (Olson 2014:158).

Olson does not elaborate on the claim he makes in this passage. This is unfortunate, since it is not entirely clear what type of claim he intends to make here. Does he intend to make a substantial claim about what kinds of arguments error theorists *ought* to offer, or what debates they *ought* to find important? I find this interpretation implausible, mainly because I fail to see how it would meet the challenge. The truth of the claim that error theorists ought to offer only arguments to the effect that error theory is true seems compatible with the truth of the claim that that no one would make a rational mistake in rejecting error theory or be epistemically praiseworthy in accepting it, since this claim is an entailment of the claim, accepted by error theorists, that there are no epistemic reasons. Does this mean that Olson intends to make a descriptive claim about what arguments error theorists *are* offering, and what debates *are* important to them? If so, his claim seems to be that error theorists are only offering arguments to the effect that moral error theory is true, and find important only debates that concern where the truth lies in metaethics and metaepistemology. But if this is the claim Olson intends to make, it is unclear whether his claim is true. It seems possible that there is some (possibly confused) error theorist who *is* offering arguments to the effect that there is epistemic reason to believe that error theory is true, and who *does* find debates concerning what there is reasons to believe important.¹⁴ Moreover, even if there is no such error theorist, this does not provide a satisfactory answer to the dialectical version of the problem, mentioned above. Even if *the error theorist* is only interested in debates concerning where the truth in metaethics and metaepistemology lies, *we* (who are not error theorists, or have not yet made up our minds as far as the truth of error theory is concerned), might consider debates on what there is reason to believe, or what would be rational to believe, important. Hence, *we* might still conclude that the error theorist's arguments are no good prior to listening to these arguments. Perhaps I am missing something here.

There is however a potentially fruitful way of answering the latter objection. The error theorist might ask us (i.e., her opponents) to consider the arguments for error theory under the assumption that there *are* epistemic reasons. If, when operating under this assumption, we arrive at the conclusion

¹⁴ Thanks to Frans Svensson for helpful comments in relation to this.

that there are no epistemic reasons, we have a problem *by our own lights*. If the error theorist is able to convince us of this, it would seem that her position is not, after all, polemically toothless.¹⁵

5.3 A challenge to the entailment claim

In the previous section, I accepted the claim that that moral error theory entails epistemic error theory for the sake of argument, and indicated some ways in which the epistemic error theorist may attempt to meet the challenges formulated by Cuneo. In this section, I challenge the entailment claim by arguing that epistemic reason can be understood as hypothetical reasons and hence as interestingly disanalogous to moral reasons. If the entailment claim is rejected, *moral* error theorists need not be worried by any problematic implications of epistemic error theory.

We saw above that the entailment claim can be motivated by the claim that the moral and epistemic domains are analogous in two ways: First, moral and epistemic *reasons* are analogous in the following way: that feature of moral reasons on the basis of which moral error theorists have argued there are no (or cannot be any) moral reasons – e.g. normative irreducibility or non-institutionality – is also a feature of *epistemic* reasons, if they exist. Second, (positive) moral and epistemic *judgments* and *statements* are analogous in the sense that while a positive moral judgment or statement is true only if there are, or can be, moral reasons, a positive *epistemic* judgment or statement is true only if there are, or can be, epistemic reasons. Consequently, so the argument goes, if the moral error theorist's arguments to the effect that there are no (or cannot be any) moral reasons are successful, epistemic error theory is true.

Below, I defend Kornblith's (1993) claim that epistemic reasons are hypothetical reasons. If this claim is plausible, the truth of (positive) epistemic judgments and statements *does not* require that there are epistemic reasons which share that feature of moral reasons on the basis of which moral error theorists have argued that there are no (or cannot be any) moral reasons. There may still be epistemic reasons. But if there are, they are not analogous to moral reasons in this particular sense. If so, moral error theory *does not* entail epistemic error theory.

Moral error theorists may be in agreement on the truth of the claim that there are no (or cannot be any) moral reasons for action (as presupposed or assumed in ordinary moral discourse) without being in agreement on in virtue of what (presupposed or assumed) feature of moral reasons this claim is true. But almost regardless of in virtue of what feature of moral reasons they find such reasons problematic, moral error theorists find *hypothetical* reasons for action – i.e., reasons to take the means to satisfy one's desires, achieve one's

¹⁵ Thanks to Gunnar Björnsson for this suggestion.

aims or goals, promote one's interests, projects, etc. – acceptable.¹⁶ (From now on, “desire” is used broadly, to cover desires, ends, goals, interests and related states, unless otherwise indicated.) We saw in chapter 3 that a hypothetical reasons claim is a claim that an agent, *S*, has a reason to take the means to satisfy his desires. More specifically:

if an agent, *S*, desires *X* and ϕ :ing is a means of achieving *X*,
then *S* has a hypothetical reason to ϕ .

In other words: In order for *S* to have a hypothetical reason to ϕ , it has to be the case *both* that *S* desires *X*, and that *X* is causally connected to ϕ . If either condition is not satisfied, it is not the case that *S* has a hypothetical reason to ϕ . In this sense, a hypothetical reason depends for its existence both on a desire and a causal connection (between the thing desired and an action that is a means of achieving the thing desired) (Joyce 2001:35, 122; Olson 2014:152-5).

It is reasonable to suppose that if error theorists can consistently accept that agents can have hypothetical reasons *for action*, they can also consistently accept that agents can have hypothetical reasons *for doxastic behavior* such as forming (certain) beliefs, revising beliefs, or abandoning beliefs. Moreover, if hypothetical reasons for doxastic behavior relate to desires for truth or coherence, or to other desires the contents of which are *epistemic* in a relevant sense, it seems reasonable to refer to these reasons as *epistemic* reasons.¹⁷ If so, error theorists can affirm that there are *hypothetical epistemic reasons* for belief (or other doxastic behavior), consistent with denying that there are normatively irreducible, non-institutional desire-transcendent, etc., reasons for belief.

Kornblith provides a forceful suggestion as to how the details of an account of hypothetical epistemic reasons might be spelled out. According to this account, all agents who desire something have hypothetical reasons to conduct her (doxastic) actions in accordance with epistemic norms. (From now on, I use the phrase “*S* follows norm *N*” as shorthand for “*S* conducts her (doxastic) actions in accordance with *N*”.) More specifically, for any agent, *S*, if *S* desires *X*, following epistemic norms is a means of achieving *X*. Consequently, *S* has a hypothetical reason to follow epistemic norms. Kornblith says that

[p]recisely because our cognitive systems are required to perform evaluations relative to our many concerns, and to perform these evaluations accurately, the standards by which

¹⁶ Joyce (2001:177) claims that his arguments for moral error theory do not threaten the existence of hypothetical imperatives. According to Olson (2014:152), many moral error theorists want to accept hypothetical reasons, and hold that hypothetical reasons claims are sometimes true.

¹⁷ That the things *avored* – cognitive states and processes – are epistemic is however *not* sufficient to render the reasons themselves epistemic, since a belief, for example, may be favored also by moral or aesthetic considerations (Kornblith 1993:368).

we evaluate these cognitive systems themselves must remain insulated from most of what we intrinsically value, whatever we may value. This provides a reason to care about the truth whatever we may otherwise care about. It also provides us with a reason to evaluate our cognitive systems by their conduciveness to truth. And this is precisely what epistemic evaluation is all about (Kornblith 1993:372).

The claim made in this passage can be illustrated as follows: Suppose that *S*, who desires *X*, stands before a collection of available courses of action $\phi_1, \phi_2, \phi_3, \dots, \phi_n$. Because *S* desires *X*, *S* wants (*ceteris paribus*) to perform that action, whichever it is, that furthers *X*. In order to do so, she needs to be able to correctly identify this action among the available alternatives. Suppose that ϕ_1 is the only action out of those available that furthers *X*. If so, *S* needs to be able to find out that this is the case in order to perform the action that furthers her desire. Because following epistemic norms increases *S*'s chances of ending up with beliefs that represent reality correctly, and having such beliefs with respect to what action actually furthers *X* is a means of achieving *X*, *S* has hypothetical reasons to follow epistemic norms.

Implicit in this illustration is the claim that whatever serves as the object of *S*'s desire, the fact that *S* desires it entails that *S* wants (*ceteris paribus*) to act in ways that further it. This seems plausible. It is in the nature of desires (and related states) that we aim to make the world match them. In other words: It is an essential part of what it is for an agent to desire something that she is in some kind of motivational or dispositional state with respect to the thing she desires. If *S* desires that *p*, then if *S* observes that *not-p* is the case, *S* is disposed or motivated to change the world so that *p* is the case (Joyce 2001:111). This supports the claim that if *S* desires *X* and stands before a collection of available courses of action, *S* (*ceteris paribus*) has a preference for that action, whichever it is, that furthers *X*.

Another claim made in this illustration is that following epistemic norms increases *S*'s chances of finding out what action(s) actually furthers *X*. This is supported by the claim that epistemic norms concern, in some way or other, the goal of correct representation of reality (for a similar claim, see Cuneo 2007:56-7), or truth. Take, for example, the norm that subjective confidence is to be apportioned in line with evidential support. This norm is related to the goal of correct representation of reality, or truth, in the sense that by applying it consistently and carefully, an agent greatly increases the probability of ending up with beliefs that are correct representations of reality, or true. In general, agents who follow norms related to the goal of correct representation of reality, or truth, will to a greater extent end up with beliefs that correctly represent reality, or are true, than agents who do not.

If this is correct, it starts to seem plausible that whatever we desire, following epistemic norms is a means of satisfying our desires, and (consequently) that whatever we desire, we have hypothetical reasons to follow epistemic norms. Kornblith says that

[...] epistemic evaluation finds its natural ground in our desires in a way which makes truth something we should care about whatever else we may value (Kornblith 1993:373).

[...] someone who cares about acting in a way which furthers the things he cares about, and that includes all of us, has pragmatic reasons to favor a cognitive system which is effective in generating truths, whether he otherwise cares about truth or not. We should thus adopt a method of cognitive evaluation which endorses truth-conducive processes (ibid:371-2).

[...] we need to make evaluations of alternative courses of action and, whatever we care about, we need these evaluations to be done accurately, i.e. by a cognitive system which generates truth (Kornblith 1993:372).

I take the second and third passages to clarify the first. Hence, I interpret Kornblith as using the phrase “something we should care about” in the first passage in a non-normative sense, i.e. as meaning “something we have practical use for” or “something we have pragmatic or instrumental reasons to favor”.

Moreover, according to Kornblith, “epistemic norms [are] a variety of hypothetical imperative [...] [and] these imperatives [are] universal (ibid:357).” He says that

on the one hand, [...] [epistemic norms] are derived from our desires in a way which removes any mystery surrounding them, and, on the other, [...] they are universal in their applicability and not merely contingent upon having certain values (ibid:372).

Note that Kornblith is not attempting to defend any particular epistemic norms. He is not claiming, that is, that there is a specific set of epistemic norms such that all agents who desire something have hypothetical reasons to follow them.¹⁸

¹⁸ Nor does Kornblith claim that there is some specific epistemic norm *N* – e.g. to apportion subjective confidence in line with evidential support – such that for any desire held by any agent, following *N* is a means of achieving it. We sometimes desire things that can be achieved only if we ignore some epistemic norm(s). If Miranda desires to believe that everything that Brad-the-3-year-old tells her is true, Miranda’s apportioning her subjective confidence in line with evidential support might be more likely to hinder than to promote satisfaction of this specific desire of hers. (More on this below.)

Instead, Kornblith seems to make the general claim that because we desire things, we typically have hypothetical reasons to care about truth and truth-conduciveness and thus have hypothetical reasons to evaluate doxastic states and actions in relation to these goals. For this reason, agents generally have hypothetical reasons to follow norms that concern the correct representation of reality – i.e., epistemic norms. Given this, the following is a minimal way of understanding Kornblith’s claim that epistemic norms are *a variety of hypothetical imperative*: If *S* desires *X*, there is some (at least one) epistemic norm, *N*, such that following *N* is a means of achieving *X*. Therefore, *S* has a hypothetical reason to follow *N*. Given a more general interpretation, Kornblith’s claim is that if *S* desires *X*, following epistemic norms is a means of achieving *X*, and that *S* therefore has a hypothetical reason to follow epistemic norms. Regardless of which interpretation we settle for, Kornblith’s claim implies that epistemic norms provide *S* with (hypothetical) epistemic reasons for belief (and other doxastic behavior) only if *S* meets a specific condition. The condition is that *S* has at least one desire directed towards something.

Let us now consider the claim that imperatives generated by epistemic norms are *universal*. Kornblith explains what he means by this in the following passage:

[...] there are certain substantive constraints which we will want our cognitive systems to meet simply on the condition that we value anything at all. Since everyone values something or other, this condition is universally met [...] (Kornblith 1993:359).

It is a reasonable empirical supposition that (at least) *close to all* agents meet the condition. It is reasonable to suppose, that is, that for almost any agent, there is at least one thing that this agent desires. If this is correct, epistemic norms provide almost all agents with hypothetical reasons. But Kornblith’s claim that imperatives generated by epistemic norms are hypothetical entails that *if* epistemic norms do provide all agents with reasons, they do not do so by necessity. Consequently, it is consistent with what Kornblith suggests that there is some agent who (at least presently) has no desires (a strongly devoted Buddhist monk or a severely depressed and heavily medicated person comes to mind). If there is such an agent, epistemic norms do not provide her with any reasons.

Kornblith’s account seems to imply that epistemic norms are universal in another sense. According to his account, recall, it is true for any agent that if she desires at least one thing, she has hypothetical reasons to follow epistemic norms. Hence, if there is an agent who has no desires, then although epistemic norms do not provide her with any reasons, they apply to her in the sense that *if* she would have desired something, she would have had reason to follow them. If this is what Kornblith claims, we can test his claim empirically by considering

some agent who has at least one desire. If Kornblith is right, we should be able to find at least one epistemic norm, *N*, such that for some of this agent's desires, following *N* is a means of satisfying it. If the agent has merely *one* desire, then there is some epistemic norm, *N*, such that following *N* is a means of satisfying *this particular desire*.

But what if there is an agent whose one and only desire concerns some feature of her own beliefs *unrelated* to truth (such as how pleasurable they are), or an agent whose one and only desire is such that *ignoring* epistemic norms is a means of satisfying it? It may appear that the existence of such agents would prove Kornblith wrong. I believe that this appearance is misleading. For illustration, I will consider one example of each kind.

To start with an example of the first kind, suppose that Peggy's one and only desire is to have beliefs such that her having them maximizes her level of pleasure. If so, Peggy has no (independent) desire to have beliefs that are true or constitute correct representations of reality, and it may appear that she has no desire such that following epistemic norms is a means of satisfying it.¹⁹ Suppose that Peggy is choosing between forming the belief that *aliens assassinated Kennedy* and forming the belief that *Oswald assassinated Kennedy*. Because her only desire is to have pleasure-maximizing beliefs, she wants to form that belief, whichever of them it is, which maximizes her pleasure. Moreover, because the truth of the matter – i.e., the fact of who assassinated Kennedy – is completely irrelevant to her choice, it appears that she has no use for principles the following of which increases the probability of her beliefs' representing reality aright, or being true. She *would have* use for principles that function to ensure that she ends up with pleasurable beliefs. But this is not the function of epistemic norms.

I believe that this appearance is misleading. I believe, that is, that it can be argued, on behalf of Kornblith's suggestion, that epistemic norms are instrumentally valuable even to someone like Peggy. While Peggy does not care about whether her beliefs are correct representations of reality, or true, she does care about whether they maximize her pleasure. When choosing between forming the belief that *aliens assassinated Kennedy*, and forming the belief that *Oswald assassinated Kennedy*, Peggy wants to perform that (doxastic) action which *actually* leads to the maximization of her pleasure-level – i.e., she wants to start believing whichever proposition the belief in which *actually* maximizes her pleasure-level. But in order to identify which one of the available actions will result in a belief that has this feature, Peggy needs to do something that resembles a *cost-benefit analysis*. Given this, Peggy's desire is best served by her having a cognitive system that reliably produces correct representations of reality (or truths), or in other words, by her following epistemic norms. This being said, it does seem possible for Peggy to satisfy her desire to hold beliefs

¹⁹ Thanks to Ragnar Francén Olinder for drawing my attention to this challenge.

that maximize her pleasure-level by accident. I am not denying that this could happen without Peggy's having any true beliefs. But it seems plausible that in order for her to *non-accidentally* satisfy this desire, she needs true beliefs about what *actually* maximizes her pleasure level.

I take this example to illustrate that when Peggy wants to evaluate different courses of action, whether of the doxastic kind or not, then if she has previously made revisions in her cognitive system which makes it impossible, or at least harder, for this system to generate truths or correct representations of reality, she will have done herself a disservice *by her own lights* (Kornblith 1993:372). In order for such revisions to constitute a disservice to Peggy by her own lights, it is not necessary for her to have an independent desire to believe whatever is true, to believe only true propositions, etc. It is sufficient that she has some desire directed towards *something*, because it is in the nature of desires that if an agent is in such a state, she wants to act in ways that further the satisfaction of her desire (rather than frustrate it). Even if Peggy's only desire is (e.g.) to maximize her level of pleasure, she needs her cognitive apparatus to generate correct representations of reality, or truths. If her cognitive apparatus lets through false beliefs, or incorrect representations of reality, she will not be able to determine what action (whether doxastic or not) will actually bring her maximal pleasure. If so, she will not be able to make an informed choice concerning which action best satisfies her desire to have pleasurable beliefs. She will, in other words, risk performing an action that is suboptimal as far as the satisfaction of this desire is concerned.

Let us now consider an example of an agent whose one and only desire is such that *ignoring* epistemic norms is a means of satisfying it. This example is supplied by Cuneo:

Suppose [...] [that] Ella intrinsically desires only psychological contentment. Suppose also [...] [that] she is extraordinarily psychologically fragile; for her to be psychologically content implies that she must isolate herself from most others and engage in very simple tasks. Most importantly, she must 'wall-off' from consideration a huge number of propositions about her own psychological state, the general state of the world, the status of her family and friends, and so on, as she finds considering any of these subject matters profoundly disturbing. Imagine, finally, she is very effective at doing this over the long run (Cuneo 2007:207).

Cuneo interprets Kornblith as accepting the following general principle:

- (C) [I]f a person intrinsically desires to Φ , and Ψ ing is, over the long run, the most likely way to satisfy her intrinsic desire to Φ

(and she is capable of desiring to Ψ), then she should desire to Ψ (ibid:206).²⁰

According to Cuneo (ibid:207), (C) implies that Ella *has a reason to desire* not to form any true beliefs with respect to propositions that concern her own psychological state (etc.), and that this implies that epistemic norms are *not binding* for her.

I understand Cuneo's argument as follows: Because the most likely way for Ella to satisfy her intrinsic desire is for her to avoid considering propositions about her own psychological state (etc.), (C) entails that Ella has a reason to desire not to consider these propositions. Presumably, forming a belief about some proposition involves considering it. Consequently, if Ella has a reason to desire not to consider propositions about her own psychological state, (C) entails that she has a reason to desire not to form any true beliefs about them either. Moreover, if following epistemic norms greatly increases the probability of ending up with true beliefs (and ignoring epistemic norms greatly decreases the probability of ending up with true beliefs), then the fact that the most likely way for Ella to satisfy her intrinsic desire is for her *not* to form true beliefs about propositions that concern her own psychological state seems to suggest that Ella has a reason to desire *not to follow epistemic norms* with respect to these propositions. Because epistemic norms are *binding* for Ella only if conforming to them offers Ella her best chance of satisfying her desires, epistemic norms are not binding for Ella with respect to these particular propositions. Cuneo concludes that

[t]he general principle to which Kornblith's argument appeals [...] [viz. (C)] can be used equally well to show that, for a large domain of propositions, some agents have reasons *not* to care about believing them to be true. [...] If this is right, [...] [Kornblith's argument] does not imply that there are epistemic norms that are universally binding. For it does not rule out

²⁰ This highlights two differences between mine and Cuneo's interpretations of Kornblith: First, Cuneo interprets Kornblith as claiming that whether *S* has reasons to follow epistemic norms depends on *S*'s *intrinsic* desires, rather than more loosely on *S*'s desires and related states. I believe that the looser interpretation squares better with Kornblith's alternation between describing epistemic norms as "contingent upon having any goals whatever (Kornblith 1993:364)", "conditional upon having any goals at all (ibid)", "grounded in desire (ibid:359)" and "derived from our desires (ibid:372)." Second, while I interpret Kornblith as claiming that if *S* desires *X*, *S* has *hypothetical reasons* (e.g.) to apportion subjective confidence in line with evidential support, because doing so will greatly increase her chances of identifying what action is actually a means of satisfying her desire, Cuneo interprets Kornblith as claiming that if *S* (intrinsically) desires *X*, *S* has *reasons to desire* that she (e.g.) apportions subjective confidence in line with evidential support, because desiring to do so is (in the long run) the most likely way to satisfy her desire. According to this account, it is not *S*'s *following epistemic norms* that does the trick, but rather *S*'s *having a desire to do so*. Part of the reason why our interpretations differ might be that Kornblith sometimes leaves his points somewhat open to interpretation. These interpretative differences do not seem to affect the force of my objection to Cuneo, below.

cases in which (i) two agents are exactly mentally alike apart from the fact that S lacks certain intrinsic desires (ii) both are aware that a proposition *p* is decisively supported by the evidence, but (iii) only one agent has an epistemic reason to believe *p*. At most, Kornblith's argument implies that, since every agent has intrinsic desires, every agent has reason to conform to epistemic norms of some sort or other, however idiosyncratic they may be (Cuneo 2007:207-8).

I agree with Cuneo that Kornblith is committed to the claim that for a specific set of propositions – those that concern her psychological state – Ella has *no reason* to (desire to) follow epistemic norms, and possibly even a reason to (desire to) *not* follow epistemic norms. However, because I do not believe that this implies that Ella does not have practical use for true beliefs, or beliefs that represent reality aright, *on some matters*, I do not believe that this implies that Ella does not have practical or instrumental reasons to favor a cognitive system which is effective in generating truths or correct representations of reality *on those matters*. In short, the fact that Kornblith is committed to this claim does not entail that Ella does not have hypothetical reasons to follow epistemic norms.

First, it is possible that Ella *has* reason to care about the truth also of the walled off propositions. Although psychological contentment is her only intrinsic desire, there are presumably also other things she desires, and having true beliefs with respect to propositions about her own psychological state (etc.) may be instrumental to satisfying these desires. (It may of course be the case that because of her fragility, these reasons are comparatively weak, and that she has stronger overall reason not to consider the propositions in question.)

Second, I believe that in analogy to how Peggy would be able to (non-accidentally) satisfy her desire to hold beliefs that maximize her pleasure level only if she had true beliefs about what beliefs *actually* maximize her pleasure level, Ella will be able to (non-accidentally) satisfy her desire to be psychologically content only if she has true beliefs about how propositions about her own psychological state are *actually* best “walled of” from consideration, and how true beliefs about these propositions are *actually* best avoided. Supposedly, there are several things Ella needs to do, and avoid doing, in order to avoid considering propositions that concern her own psychological state. If she gets letters sent to her from the hospital, reading them will greatly increase the risk of her considering her own psychological state (and also of her forming true beliefs about this matter). So she needs to take precautions to avoid reading any such mail, perhaps by asking someone else to secretly go through her mail every day and remove any letters from the hospital. If she has appointments at the hospital's psychiatric ward, this will greatly increase the risk of her considering her own psychological state. So she needs to take measures to avoid

understanding that she has these appointments, or what type of place the psychiatric ward is and for what reasons people get appointments there. She might achieve this by taking some mind-altering drug before visiting the hospital, by blindfolding herself, etc. The details are not important. The important point is that if Ella needs to avoid considering propositions about her psychological state in order to (non-accidentally) satisfy her desire to be psychologically content, this desire can contribute to its own satisfaction *in normal ways* only by working in tandem with (i.e. guide her behavior in combination with) *suitable true beliefs*.²¹ In short: Ella needs true beliefs for her desires to lead to their own satisfaction in a normal way. So it seems that Ella has hypothetical reasons to follow epistemic norms: there is something Ella desires such that in order to (non-accidentally) achieve it, she needs a cognitive system which is effective in generating truths or correct representations of reality.

Third, I disagree with Cuneo's claim that Kornblith has failed to show that epistemic norms are universally binding, because his theory allows for cases where

[...] (i) two agents are exactly mentally alike apart from the fact that *S* lacks certain intrinsic desires (ii) both are aware that a proposition *p* is decisively supported by the evidence, but (iii) only one agent has an epistemic reason to believe *p* (Cuneo 2007:208).

The reason why I disagree with this is that I do not believe that a case where (i)-(iii) are true is a case where epistemic norms are not universally binding. Rather, this seems to be a case where epistemic norms are in some sense *context-dependent*: they provide agents with reasons in different extents. They nevertheless supply both agents with reasons in some non-negligible extent, i.e. with respect to some doxastic actions and some propositions. This is precisely what I take the examples with Peggy and Ella to suggest. Kornblith *is* indeed committed to the claim that if two agents, *S*₁ and *S*₂, have different desires, it is possible that there is some context where *S*₁ has hypothetical reasons to perform some doxastic action ϕ , while *S*₂ does not have hypothetical reasons to ϕ .²² But this does not entail that *S*₂ has no hypothetical reason *at all* to follow epistemic norms. I have argued that it is reasonable to suppose that *S*₂ has some desire such that in order for this desire to lead to its own satisfaction in a normal way, *S*₂ needs beliefs that are true or represent reality aright. Consequently, it is reasonable to suppose that there are *some* propositions (e.g. those that do not

²¹ This is strictly speaking not entirely true, because satisfaction of desires is also ensured in normal *but less common* ways by expressions of desire plus helpful others doing the relevant means-end reasoning. Thanks to Gunnar Björnsson for pointing this out.

²² Note that *S*₂ might still have *moral* or other *non-epistemic* reasons to ϕ .

concern S_2 's own psychological state), and *some* doxastic actions with respect to which S_2 has pragmatic or instrumental reasons to follow epistemic norms. It is reasonable to suppose, that is, that there is *some* context where S_2 has hypothetical reasons to follow epistemic norms. Moreover, I have claimed that this seems to suggest that S_2 has pragmatic reasons to act according to epistemic norms at all times, or to form the habit of doing so, because in any context C , S_2 's chances of (non-accidentally) satisfying her desires will decrease if she has not followed these norms in contexts temporarily preceding C .

Fourth, Cuneo claims that Kornblith has at most shown that "since every agent has intrinsic desires, every agent has reason to conform to epistemic norms of some sort or other, however idiosyncratic they may be (Cuneo 2007:208)." Even granting that this is true, this does not seem to pose any forceful challenge to the Kornblith's account of epistemic normativity as spelled out above. As I understand Kornblith, his point is not that there is a specific set of epistemic norms such that all agents who desire something have hypothetical reasons to follow them, but rather that because we desire things, we typically have hypothetical reasons to care about truth and truth-conduciveness and thus have hypothetical reasons to evaluate doxastic states and actions in relation to these goals. For this reason, agents typically have hypothetical reasons to follow epistemic norms – norms that are related to these goals. This general point is consistent with it being the case that in some contexts, we have no reason to care about truth, or have reasons not to care about truth.²³

5.4 Conclusion

This chapter has discussed what I have referred to as *the epistemic objection* to moral error theory, and raised two challenges to it.

First, I indicated some ways in which the epistemic error theorist (or the moral error theorist who accepts the entailment claim) can attempt to meet the challenges to epistemic error theory formulated by Terence Cuneo.

Second, I challenged the entailment claim by giving an account of Hilary Kornblith's suggestion that epistemic reasons are hypothetical reasons. I argued that most of us have hypothetical reasons to follow epistemic norms. The argument for the entailment claim presupposes that moral and epistemic reasons are analogous in the sense that the feature, *f*, of moral reasons on the basis of which moral error theorists argue there are no (or cannot be any) moral reasons – e.g. normative irreducibility or non-institutionality – is also a feature of epistemic reasons, if they exist. But moral error theorists generally do not believe that hypothetical reasons have *f*. If epistemic reasons are hypothetical, moral error theorists can reject both the claim that moral and epistemic reasons

²³ For discussion of why epistemic norms could be subject to a Kornblith-style treatment, whereas moral norms could not (or at least not as easily or obviously), see chapter 5.

are analogous in the sense presupposed by the argument for the entailment claim, and the claim that their arguments for the claim that there are no (or cannot be any) moral reasons, if successful, establish the truth also of epistemic error theory.

If these arguments are sound, there are ways for the moral error theorist to meet at least some of the challenges faced by epistemic error theory, and good reasons for her to reject the claim that moral error theory entails epistemic error theory. If so, the epistemic objection does not convincingly show that moral error theory has problematic implications.

Chapter 6

The psychological objection

The topic of this chapter is what I refer to as *the psychological objection* to ERROR THEORY. According to this objection, moral error theory implies error theory with respect to all normativity – *normative error theory*. Normative error theory is claimed to have the problematic consequence that it cannot be believed. Consequently, if the psychological objection is successful, moral error theory has entailments that cannot be believed. Moreover, the unbelievability claim appears to entail *error theory with respect to thoughts and discourse about beliefs in the normative error theory*: it appears that if it is true, beliefs and claims to the effect that someone believes normative error theory are systematically in error.

Bart Streumer (2013, 2016a, 2016b) gives an argument for unbelievability claim. In defending this claim, Streumer does not intend to defend the psychological objection, nor to raise any other objection to normative error theory. On the contrary, he suggests that while our inability to believe normative error does not undermine its truth, it does undermine many objections to it (Streumer 2013:195, 201 ff.).¹ Streumer’s argument for the unbelievability claim goes as follows: If normative error theory is true, there are no normative properties, and therefore no reasons to believe normative error theory. We cannot fail to believe what we believe to be entailed by one of our own beliefs, and cannot have a belief while believing that there is no reason for this belief (Streumer 2013:197-8).

I challenge the psychological objection by challenging Streumer’s argument for the unbelievability claim. First, I suggest that even if we cannot have any normative reasons to believe normative error theory, we typically have hypothetical reasons to follow epistemic norms, and may for this reason have hypothetical reasons to believe normative error theory. Second, I suggest that when “belief” is understood in a way that makes the premises of his argument plausible, its conclusion is not as radical as it may first seem, and need not be taken to imply error theory with respect to belief in normative error theory.

¹ Among them, the “toothlessness challenge” discussed in chapter 5 (see Streumer (2013:203-5).

I proceed as follows: First, I give an account of how the claim that moral error theory entails normative error theory can be defended (section 6.1.). Second, I give an account of Streumer's argument for the unbelievability claim, and indicate some possible responses (sections 6.2 and 6.3).

6.1 The entailment claim

Several writers accept the claim that moral error theory entails normative error theory, in some form (from now on, "the entailment claim"). Olson says that "[a] plausible error theory must take the form of an error theory about irreducibly normative favouring relations, or more generally about irreducible normativity (Olson 2014:116)," and that "a plausible moral error theory must be an error theory about all normativity (Olson 2014:3);" Streumer (2016b:419) agrees. Matthew S. Bedke seems to implicitly accept the entailment claim when he says that "if moral reasons are metaphysically queer, all [normative] reasons are metaphysically queer (Bedke 2010:42)" and that "[t]he queerness objection applies to the favouring relations quite generally, and it does not matter wherefrom the favoring flows (ibid:54)."

The details of an argument for the entailment claim depend on the details of the claims made by the moral error theorist targeted by it. If, for example, it is directed at someone who denies the existence of irreducibly normative favoring relations, it could be formulated along the following lines: If moral error theory is true, there is no irreducible normative favoring, and if there is no irreducible normative favoring, normative error theory is true.

The psychological objection targets moral error theorists who, in one way or another, for some reason or another, hold doubts about *normativity in general*. Such doubts are not universally held by moral error theorists. Joyce, for example, notices that it is possible to be an error theorist about all normative phenomena (and also that it is in principle possible to be a radical *global* error theorist), and claims that "typically the moral error theorist thinks that there is something *especially* problematic about morality, and does not harbor the same doubts about normativity in general (Joyce 2007:52)."² Possible strategies for defending moral error theory without raising doubts about normativity in general – i.e., without becoming a target of the psychological objection – includes arguing that moral discourse presupposes that agents can have non-institutional desire-transcendent reasons for action while denying that normative discourse in general involves such a presupposition, and arguing that all metaethical views except moral error theory fail (see chapter 1 and Joyce 2007:52, 2011a:520, 2016a).³

² See also Joyce (2005:288).

³ The issue of what scope moral error theory should properly be taken to have divides error theorists into two camps: those who believe that the theory can be restricted to the metaethical domain, or at

6.2 Streumer's argument for the unbelievability claim

This section discusses Streumer's (2013, 2016a, 2016b) argument for the claim that normative error theory cannot be believed (from now on, "the unbelievability claim"), and raises two challenges to it. If normative error theory is analogous to ERROR THEORY, it entails that (positive) normative judgments and statements have truth-aspirations but uniformly fail to secure truth. According to Streumer's account, normative error theory also involves the claim that there are no normative properties (Streumer 2013:194).

Streumer's argument for the unbelievability claim goes as follows:

- (S₁) [p] If normative error theory is true, there are no normative properties,
- (S₂) [p] The property of being a reason for belief is a normative property,
- (S₃) [1, 2] If normative error theory is true, there is no reason to believe normative error theory,
- (S₄) [p] We cannot fail to believe what we believe to be entailed by one of our own beliefs,
- (S₅) [3, 4] If we believe normative error theory, we believe that there is no reason to believe normative error theory,
- (S₆) [p] We cannot have a belief while believing that there is no reason for this belief,
- (S₇) [5, 6] We cannot believe the normative error theory (ibid:197-8).

When combined with the entailment claim, the unbelievability claim suggests that moral error theory entails a claim that cannot be believed. Moreover, the unbelievability claim appears to entail *error theory with respect to thoughts and discourse about beliefs in the normative error theory*. It is reasonable to suppose that the normative error theorist believes that she believes normative error theory. It is also reasonable to suppose that if she believes this, or makes a statement to the effect that this is the case, she holds a belief or makes a statement that aims the truth. But it appears that if the unbelievability claim is true, it is not the case that anyone believes normative error theory. Consequently, it appears that if the unbelievability claim is true, beliefs and statements to the effect that someone believes normative error theory aim at the truth but systematically fail to secure it.

least to the moral domain more generally, and those who believe that it cannot be so restricted. In the latter camp, there is room for disagreement on how far the implications of the theory stretch: some might think that it entails error theory with respect to all normativity, others that it extends only to (e.g.) the epistemic domain.

In the remainder of this chapter, I discuss Streumer's argument for the unbelievability claim. First, I discuss an objection that has been raised to (S₆), and explain how Streumer can handle it. Second, I argue that the error theorist can deny premise (S₂), and consequently deny (S₃) and (S₅) which depend on (S₂). Third, I discuss an objection that has been raised to (S₄). I argue that this objection can be challenged. Fourth, I argue (in section 6.3) that when "belief" is understood in a way that makes premises (S₆) and (S₄) plausible, the conclusion of Streumer's argument is not as radical as it may first seem, and does not entail error theory with respect to thoughts and discourse about beliefs in normative error theory (from now on, "error theory" and "error theorist" and refers to *normative* error theory and normative error theorists, unless otherwise indicated).

Let us begin by examining premise (S₆) of the above argument. Streumer makes three claims relevant to correctly understanding this claim. First, he says that he uses "belief" to mean *full belief* throughout his argument. Full belief contrasts to partial belief. The difference concerns how confident the believer is in the truth of the proposition believed. If someone is wholly (or very) confident that *p*, she fully believes that *p*. If she is fairly but not wholly confident that *p*, she partly believes that *p*. Moreover, if she considers giving up the belief that *p*, she at most partially believes that *p* (Streumer 2013:195, 2016b:424). Second, Streumer says that his claim may not be true of *compulsive beliefs*. It may be possible, he says, to have a full compulsive belief that *p* while at the same time fully believing that there is no reason to believe that *p* (Streumer 2013:197). Third, he says that he uses "reason to believe" to mean "normative reason to believe" – i.e. a consideration *counting in favor* of a belief. Because the favoring relation is a normative relation, a belief that there is a reason for believing *p* is a normative judgment (i.e., a normative belief) (Streumer 2013:196).⁴

Given this, I understand (S₆) as stating that nothing could be a full, non-compulsive belief if it could co-exist with awareness that it was not based on reasons. A state that could so co-exist might at most be belief-like (where "belief" is understood as full, non-compulsive belief):

Premise (S₆) If *S* fully and non-compulsively believes that *p*, then *S* does not fully and non-compulsively believe that there is no normative reason (whether moral, epistemic, aesthetic, or what have you) to believe that *p* (see Streumer 2013:196; 2016b:426).

⁴ The class of normative reasons is usually taken to include (but need not be understood as limited to) moral, epistemic, and aesthetic reasons. Normative reasons are *in general* understood in terms of the normative "counting in favor"-relation. Regardless of their nature, the fact and the action are related as follows: the fact counts in favor of the action.

In defense of this claim, Streumer (2013:196) asks us to consider a claim of the form “ p , but there is no reason to believe that p ”, such as

- (a) Socrates was mortal, but there is no reason to believe that Socrates was mortal.

Streumer argues as follows:

[...] [S]uppose that someone says: “Socrates was mortal but there is no reason to believe that Socrates was mortal.” [...] [T]his person may be insincere, or may fail to understand what he is saying, or may be considering whether to give up one of these beliefs. If so, he does not fully believe what he says he believes. Alternatively, he may be sincere, may understand what he is saying, and may not be considering whether to give up one of these beliefs. But if so, he is too confused to fully believe what he says he believes (Streumer 2013:196).

If either of these are true descriptions of an agent, S , it is not the case that S has a full non-compulsive belief that Socrates was mortal and that there is no reason to believe that Socrates was mortal. It might be that S believes at least one of the conjuncts compulsively, or believes at least one of them only partially – i.e. that with respect to at least one of the conjuncts, S believes it without full confidence, or is considering revising or abandoning it. It might also be that S holds some *other* propositional attitude with the content that Socrates was mortal and there is no reason to believe that Socrates was mortal. Perhaps she *desires*, *wishes*, or *hopes* it. It is also possible that S is *pretending* or *role-playing* that Socrates was mortal and that there is no reason to believe that Socrates was mortal (or takes this attitude to merely one of the conjuncts).⁵ Either way, S does not *fully and non-compulsively believe* that Socrates was mortal and that there is no reason to believe that Socrates was mortal.

One way in which the error theorist may attempt to challenge premise (S_6) is by objecting, with Hallvard Lillehammer and Niklas Möller (2015:455), that because the kind of belief at issue is *full* belief, and because the fullness of a belief is a measure of the agent’s *degree of confidence* with respect to the truth of the proposition believed, the question of whether she is too confused to believe what she is saying is irrelevant to the issue of whether she fully believes the proposition in question. In other words: If S is sincere and understands the claim she is making, it is possible for her to fully believe the proposition in question *despite being confused*.

Streumer can be understood as addressing this objection to his argument for (S_6) when he adds the following condition on full, non-compulsive belief:

⁵ For more on pretending and role-playing, see chapter 8.

- (S₈) If someone, *S*, fully and non-compulsively believes that *p*, then *S* adequately understands *p* (Streumer 2016b:424).

He also claims that

[...] there is a distinction [...] between believing *that a theory is true* and believing *the theory itself*. If we do not understand a theory, we can believe that this theory is true, but we do not thereby believe this theory itself (ibid:425).

Given this, (S₈) states a condition on full, non-compulsive belief in a theory or proposition *itself*, and entails that it is impossible for someone who does not adequately understand error theory to fully and non-compulsively believe the error theory itself. Someone who does not adequately understand error theory may at most fully and non-compulsively believe *that the error theory is true*. Perhaps she starts believing this after a respected philosophy professor tells her that this is so. But given (S₈), this would not be an instance of belief in the error theory itself (Streumer 2016b:425). Given this, Streumer seems to understand the nature of full, non-compulsive belief in the error theory (itself) in a way that entails that, contrary to what Lillehammer and Möller suggest, whether the agent is confused – i.e., does not adequately understand error theory – is relevant to whether she fully and non-compulsively believes it.

Several questions could be raised in connection to the issue of what it is for some to adequately understand error theory. Most acutely, one might wonder what criteria agents have to satisfy in order for their understanding of a proposition or theory to count as “adequate.” I return to this issue below. One might also wonder what Streumer means by “understands”. As conventionally used, for someone to understand that *p*, it has to be the case at least that she *believes* that *p* and that *p is true*. Consequently: as conventionally used, claims of the form “(agent) *S* understands that *p*” seem closely related to claims of the form “*S* knows that *p*.”⁶ But it seems implausible that Streumer uses “understands” in this way. If he did, his claim that full, non-compulsive belief entails understanding would seem to suggest that such belief entails truth.

Moreover, if (S₈) is compatible with the suggestion that an agent, *S*, who does *not* adequately understand *p* has a full non-compulsive belief that *p is true*, this seems to raise the question of what the difference is between a belief that *p* and a belief that *p is true*. Given that *p is true* entails that *p*, it would seem that if *S* fully and non-compulsively believes that *p is true* (where such belief does not entail adequate understanding), *S* is in this same state also with respect to *p itself*.

⁶ This common sense analysis of the nature of understanding can be questioned. Finnur Dellsén (2017) argues that understanding requires neither belief nor justification.

I agree with Lillehammer and Möller that it seems possible for someone to believe error theory with full (or very high) confidence even if she does not adequately understand it. Because Streumer's argument is restricted to full belief, where such belief entails adequate understanding, he may acknowledge that it is possible to believe the error theory, when belief is understood in the less restricted way Lillehammer and Möller have in mind.

Contrary to Streumer's suggestion, I believe that the error theorist may reject his claim that if someone makes a claim like (a), it is an indication that she is either insincere, does not understand the claim she is making, is not wholly confident of the truth of the claim, is considering giving up at least one of the conjuncts, or is too confused to fully and non-compulsively believe it. But the reason why this is so is that premise (S₂) is false. Streumer says that:

Premise (S₂) The property of being a reason for belief, in the sense of a consideration that counts in favor of a belief, is a normative property (Streumer 2013:197).

As understood here, error theory entails that there are no normative properties (Streumer 2013:194). Consequently, if (S₂) and error theory are both true, there are no reasons for belief, and in particular no reason to believe error theory.

The error theorist may challenge premise (S₂) is by arguing, with Olson, that although there are no (irreducibly) normative reasons for belief, "there are other senses of 'reason' in which it might well be true that there are reasons for some agents to believe certain propositions (Olson 2014:158)." The error theorist might, for example, argue that there are *hypothetical reasons to believe* that Socrates was mortal, and that this does not entail that there are normative reasons to believe that Socrates was mortal. In chapter 5, above, I defended an account of reasons for belief according to which we typically have hypothetical reasons to follow epistemic norms. If such an account of reasons for belief is successful, the error theorist might sincerely, with full confidence and adequate understanding, and without considering giving up one of her beliefs, claim

Socrates was mortal, but *there is no normative reason of any kind* to believe that Socrates was mortal.

Streumer addresses also this type of reply. In response to Olson's claim that there are senses of "reason to believe" on which agents can have reasons to believe that *p* without having (irreducibly) normative reasons to believe that *p*, Streumer says that

[...] these are odd moves for an error theorist to make. If claims about reasons for belief can be equivalent to descriptive claims

[...], why can moral claims not similarly be equivalent to descriptive claims? (Streumer 2016b:423)

Why, asks Streumer (ibid), can the moral claim that

- (b) Agent *S* morally ought to do φ ,

not be equivalent to the descriptive claim that

- (c) *S* has a desire that would be fulfilled, or is more likely to be fulfilled, if *S* does φ .

This is a legitimate question, and it may be taken to apply also to the claim that (moral error theorists may escape the claim that moral error theory entails *epistemic* error theory by arguing that) agents generally have hypothetical reasons to follow epistemic norms. I argued, in chapter 5, that because we desire things, we typically have hypothetical (or instrumental) reasons to care about truth and truth-conduciveness, and to evaluate doxastic states and actions in relation to these goals. Because epistemic norms are related to the goals of truth and truth-conduciveness, and because we need true beliefs for our desires to lead to their own satisfaction in a normal way, following epistemic norms is typically a means for fulfilling our desires, whatever we desire. Consequently, we typically have hypothetical reasons to follow epistemic norms. This claim raises the question of whether an analogous argument can be formulated with respect to moral norms – i.e., whether it can be argued that we, for similar reasons, typically have hypothetical reasons to follow moral norms.

I am ready to admit that *if* moral norms were in some analogous way related to some goal or goals that we have hypothetical (or instrumental) reasons to care about as soon as we desire something, we would typically have hypothetical reasons to follow moral norms. But moral norms are not like this: agents do not necessarily have hypothetical reasons to follow moral norms.

It is possible that for most agents, following moral norms is a means of achieving something they desire. Mackie, for example, claims that morality is a device for counteracting limited sympathies (Mackie 1977:108), and conflict-resolution (Mackie 1980:156).⁷ Given this, and that most agents as a matter of fact (individually) desire that limited human sympathies are counteracted (and so on), following moral norms is a means of achieving something they desire. But it seems improbable that *all* agents hold such desires. Consider, for example, someone like Gyges (discussed in chapter 3). Supposing that Gyges cares nothing for anyone but himself, that he only desires to rape, kill and become king, and that he uses his ring of invisibility in a way that removes all

⁷ For more on these issues, see chapter 7.

risk of getting caught, it is improbable that his following moral norms is a means of achieving anything he desires. However, even for Gyges, following *epistemic* norms is a means of achieving something he desires. Just like other agents, Gyges needs beliefs that are true (or correct representations of reality), at least on some matters, in order for his desires to lead to their own satisfaction in a normal way. Given this, moral norms differ from epistemic norms in that they are not related to some goal(s) that we have hypothetical (or instrumental) reasons to care about as soon as we desire something, *whatever we desire*.

Someone might argue, in response to this, that we have hypothetical reasons to follow moral norms because following them is a means of achieving things we desire *over and above* the specific goals to which moral norms are (most directly) related. Joyce, for example, suggests that following moral norms favors the achievement of our individual long-term desires, because doing so helps us escape weakness of will, passion, negligence, and other things that hinder the achievement of such desires (Joyce 2001:181, 184, 210 ff., 2005:300-2). Given this, and that most of agents have long-term desires, following moral norms is a means of achieving something they desire. However, this does not entail we have hypothetical reasons to follow moral norms. First, it seems that when following moral norms is a means for achieving our long term desires, it is so in virtue of constraining our choices of action, i.e. in virtue of being a means of *frustrating* some of our other desires – viz. our short-term desires. Second, it seems probable that there are agents who lack long-term desires. Consider, for example, an agent whose one and only desire is an *egoistic short-term desire* – e.g., an agent whose one and only desire is that she *herself* feels pleasure *in the near future* (a deeply nihilistic hedonist comes to mind). For such an agent, following moral norms would not be a means of achieving anything she desires. But even for this agent, following *epistemic* norms would be a means of achieving something she desires (for the same reasons as Gyges). Consequently, it seems that even if following moral norms is a means for achieving something most of us desire apart from the goals to which moral norms are related, it is not the case that following moral norms is a means of achieving what we desire, *whatever we desire*.

A third and final reason for thinking that moral claims like

- (b) Agent *S* morally ought to do φ ,

are *not* equivalent to descriptive hypothetical imperative claims like

- (c) *S* has a desire that would be fulfilled, or is more likely to be fulfilled, if *S* does φ .

is that we think of, and use, claims of these two kinds differently. As we saw in chapter 3, error theorists take moral discourse to presuppose that moral requirements are (normally) categorical in nature.⁸ That is; when a claim like

(d) Don't punch Jessica.

is used to express a moral requirement, it (normally) specifies something the addressee ought to refrain from doing *regardless of what she desires*. Suppose that (d) had been equivalent to some descriptive claim like (c), for example:

(e) Don't punch Jessica if you don't want to be punched back.

Suppose that the speaker finds out that you (the addressee) lack the relevant desire, i.e., you *do* desire to be punched back. (Perhaps you stand to win a bet if you get Jessica to punch you, or perhaps you just enjoy being punched.) If so, the speaker would normally withdraw the requirement. This is not how we normally think of moral requirements. If moral claims like (b) or (d) were equivalent to hypothetical imperative claims like (c) or (e), we would need some explanation of why we normally think that what the agent desires is irrelevant to whether she is morally required to do something, and why we normally use categorical imperatives to express moral requirements. For this reason, it seems to me that there are good reasons to reject the suggestion that moral claims are equivalent to hypothetical imperative claims.

To sum up: If moral norms were related to some goal, or goals, that we have hypothetical (or instrumental) reasons to care about as soon as we desire something, we would typically have hypothetical reasons to follow moral norms. But it seems that moral norms are *not* related to any such goal or goals. Moreover, if our way of using and thinking about moral claims could be given a straightforward explanation on the assumption that moral claims are equivalent to hypothetical imperative claims, this would speak in favor of the suggestion that moral claims are equivalent to descriptive claims of this kind. But it seems that our way of using and thinking of moral claims fits poorly with that assumption.

Given this, it seems to me that the error theorist who argues that (S₂) is false because there are senses of "reason to believe" given which the property of being a reason for belief is not a normative property may escape Streumer's challenge. There is a good answer available to her as to why, given that (some) claims to the effect that an agent has a reason to believe something are equivalent to descriptive claims, moral claims are *not* similarly equivalent to descriptive claims. And given such a sense of "reason to believe", it is possible to

⁸ We saw in chapters 3 and 4 that this is in general *not* the feature of moral requirements on the basis of which moral error theorists formulate their arguments for moral error theory.

fully and non-compulsively believe (where such belief implies adequate understanding of the propositions believed) both that p and that there is no normative reason to believe that p .

If the error theorist rejects premise (S₂), she can reject (S₃), which is entailed by the combination of (S₁) and (S₂), as well as premise (S₅), which is entailed by the combination of (S₃) and (S₄). But even if the error theorist does not reject (S₂), she may reject (S₅) by rejecting (S₄), to which I will now turn. Streumer says that:

We cannot fail to believe what we believe to be entailed by our own beliefs (Streumer 2013:195).

A person believes that p only if this person believes what he or she believes to be entailed by p (Streumer 2016b:425).

When interpreting these passages, we should keep the following in mind: First, Streumer uses “belief” to mean full belief (see Streumer 2013:195). Second, he takes full belief to imply adequate understating of the proposition believed (see Streumer 2016b:424). Third, because (S₆) was interpreted as restricted to non-compulsive beliefs, it makes most sense to interpret (S₄) as also being so restricted. Fourth, as illustrated by Marianna Bergamaschi Ganapini, the claim Streumer makes in these passages must be restricted specifically to *explicit beliefs* – i.e., belief of such a kind that the believer does not have to infer them from other beliefs – to be plausible. Bergamaschi Ganapini asks us to consider a case of the following kind:

I have the belief that p , the belief that *if p then q* , and the belief that those beliefs combined entail the belief that q . I thereby also have the belief that q . However, I might have the *explicit* belief that p and the *explicit* belief *if p then q* while at the same time having only the *implicit* belief that those beliefs combined entail the belief that q . In this case it would be possible to fail to have the *explicit* belief that q (see Bergamaschi Ganapini 2016:526).

She then points out that

This applies to the error theory as well: I may believe various claims that compose the error theory while also fail to bring some of those beliefs together in a way that it would make it obvious what they entail. As a result, I may believe the error theory but fail to realize some of its entailments (Bergamaschi Ganapini 2016:526).

In response to this, Streumer says that he uses the term “belief” to mean explicit belief.⁹ Consequently, premise (S₄) should be understood as the claim that

We cannot fail to fully and *explicitly* believe what we fully and *explicitly* believe to be entailed by one of our own full and *explicit* beliefs (Streumer 2016a:538, my italics).¹⁰

I understand premise (S₄) as stating a restriction on full, non-compulsive, explicit belief (where such belief is understood to imply adequate understanding of the proposition believed), namely that nothing could be such a belief that *p* if it could co-exist with a full, non-compulsive, explicit belief that *p* entails *q* without co-existing with a full, non-compulsive, explicit belief that *q*. A state that could so co-exist might at most be belief-like (where “belief” is understood as full, non-compulsive, explicit belief):

Premise (S₄) If *S* fully and non-compulsively explicitly believes that *p*, then if *S* fully and non-compulsively explicitly believes that *p* entails *q*, *S* fully and non-compulsively explicitly believes that *q*.

I agree with Bergamaschi Ganapini that it seems possible to believe the claims that compose the error theory while failing to realize some of its entailments. Because Streumer’s argument is restricted to explicit belief in error theory, where such belief entails explicit belief in all propositions one explicitly believes to be entailed by it, its success does not rule out the possibility of believing the error theory in this sense. This is something that Streumer may acknowledge.

Premise (S₄) seems to have implications for the interpretation of Streumer’s claim (S₈), discussed above – viz. that an agent, *S*, fully and non-compulsively believes a proposition *p* itself only if *S* adequately understands *p*. To illustrate (S₄), Streumer asks us to consider someone who makes a claim of the following kind:

⁹ This means that premise (S₆) should also be understood as stating a condition on full, non-compulsive, explicit belief. The introduction of this qualification does not affect my above discussion of premise (S₆), since it did not depend on interpreting premise (S₆) a condition on full, non-compulsive, *implicit* beliefs.

¹⁰ Streumer’s way of distinguishing between explicit and implicit beliefs makes me wonder whether he uses “explicit belief” to mean explicit *occurrent* belief. He says that “[...] we explicitly believe that *p* if we currently think that *p*, and we implicitly believe that *p* if our current thoughts commit us to *p*, for example, by presupposing or entailing *p* (Streumer 2016a:538).” In a footnote, he adds: “Another [distinction] is between *occurrent* and *dispositional* belief: we have an *occurrent* belief that *p* if we currently think that *p*, and we have a *dispositional* belief that *p* if we are disposed to think that *p* in certain circumstances, for example, if someone asks us whether it is the case that *p*. [...] I here take explicit belief to be identical to *occurrent* belief, and Bergamaschi Ganapini calls both implicit and dispositional belief ‘implicit belief’. This does not matter here (ibid:544-5).” Given this, it seems to me that it is most charitable to read Streumer as using “explicit belief” to mean “explicit *occurrent* belief”. However, because he says that it does not matter here, I will ignore this complication.

I believe that p , and I believe that p entails that q , but I do not believe that q (see Streumer 2016b:426).

Streumer says that if this person is “neither insincere nor considering whether to give up one of these beliefs, [...] he is too confused to adequately understand what he is saying, which means that he fails to meet [...] [the condition that full, non-compulsive belief implies adequate understanding] (ibid:426).” This (at least) partially answers the question, posed above, what adequate understanding amounts to: if S adequately understands p , S believes that which S believes to be entailed by it (where “belief” is understood as full, non-compulsive explicit belief). Given this, it might perhaps be argued that the difference between a belief that p , held by an agent who adequately understands p and a belief that p is true, held by an agent who does not adequately understand p is that only the former agent believes the entailments of his belief. Perhaps, someone who does not adequately understand p does not realize that her belief that p is true entails that p ?

It may seem that the plausibility of premise (S_4) varies with the contents of p and q . Olson discusses the so-called *preface paradox* in response to premise (S_4). He presents the paradox as follows:

- (i) I believe about each sentence in this book that it contains no typographical error, since I have gone over the text many times and since colleagues and reviewers have read it carefully.
- (ii) I believe that if it is true about each sentence in the book that it contains no errors, there are no errors in the book.
- (iii) However, I do not believe that there are no errors in the book. Since books without typographical errors are extremely rare, I believe that this one does contain them (Olson 2014:169-170).

While the preface paradox might show that it can be rational not to believe the consequences of everything one believes, it is not clear that it constitutes a counterexample to premise (S_4) of Streumer’s argument. The reason for this is that while the agent in the example – let’s call him Chad – believes of each sentence in the book that it contains no error, he does not believe all sentences in the book contains no error. In other words, Chad does not believe *the conjunction*.

To see this, recall that what premise (S_4) says is specifically that nothing could be a full, non-compulsive, explicit belief (where such belief is understood to imply adequate understanding of the proposition believed) that p if it could co-exist with such a belief that p entails q without co-existing with such a belief that q . Supposing that “believe” in the (i)-(iii) formulation picks out full, non-compulsive explicit beliefs, it might seem that (i)-(iii) constitutes a counterexample to (S_4) in virtue of describing an agent who fully,

non-compulsively and explicitly believes that p and that p entails q without fully, non-compulsively and explicitly believing that q . It might seem, that is, that the following is true of Chad (where “believes” picks out full, non-compulsive, explicit belief):

- (iv) Chad believes that all of the sentences $s_1, s_3, s_3, \dots, s_n$ contains no error.
- (v) Chad believes that sentences $s_1, s_3, s_3, \dots, s_n$ are all the sentences in the book.
- (vi) Chad believes that if all of the sentences $s_1, s_3, s_3, \dots, s_n$ contains no error, and $s_1, s_3, s_3, \dots, s_n$ are all the sentences in the book, there are no errors in the book.
- (vii) Chad *does not* believe that there are no errors in the book.

However, what (i) says is merely that

- (viii) Chad believes, for each of the sentences $s_1, s_3, s_3, \dots, s_n$ that it contains no error.

The reason why it is not clear that (i)-(iii) constitutes a counterexample to premise (S_4) of Streumer’s argument is that it can be argued that (viii) does not imply (iv). That Chad believes *for each sentence* that it contains no error, need not be taken to imply that he believes that *all* of the sentences contains no error (or in other words, that none of the sentences contains any error). Given this, it can be denied that Chad believes (iv). Consequently, it could be argued that Chad is not an example of an agent who fully, non-compulsively and explicitly believes that p and that p entails q without holding such a belief that q , for Chad does not believe p .

Even assuming that (viii) implies (iv), there is – as Olson (2014:170) also points out – a pretty straightforward response available to Streumer, namely that premise (S_4) concerns the nature of *full, non-compulsive, explicit belief*. Consequently, Streumer may argue that if Chad has clearly in mind his fallibility as a proof reader, he does not believe with full confidence about each sentence in the book that it contains no error. Having this fallibility clearly in mind gives Chad reason to *at least marginally* reduce his confidence, for each of the individual sentences, that it contains no error (i.e., to reduce the fullness of each of these beliefs). Consequently, it is not the case both that Chad fully believes that all of the sentences in the book contains no error, and that Chad fully believes that the book contains errors. At most, Chad holds these beliefs *to some lower degree*. In reply to this, Olson says that

It is not clear how and why this [reply] helps, however. Few defenders of error theory are as certain that the error theory is

true as they are that $2+2=4$ is true, or that p or not- p is true. Presumably, therefore, full belief should in this context not be understood as belief to degree 1, but belief to a high degree (whatever degree that is exactly). But then the preface paradox still stands. I can believe to a high degree about each sentence in this book that it contains no typographical errors and at the same time fail to believe that there are no typographical errors in the book (Olson 2014:170).

I agree with Olson that it is reasonable to suppose that error theorists in general do not believe error theory with the highest possible confidence, but merely believe it to some lower (although relatively high) degree. Because Streumer's argument is restricted to belief in error theory where such belief entails that the believer is wholly (or very) confident, its success does not rule out such belief in error theory.

This section has discussed Streumer's argument for the claim that we cannot believe the (normative) error theory. Streumer uses "belief" to mean "full, non-compulsive explicit belief", where such belief entails adequate understanding on part of the believer. Moreover, the unbelievability-claim only concerns beliefs the object of which is the error theory itself. Because the unbelievability-claim is restricted in this way, Streumer may acknowledge that it is possible for the error theorist to believe the error theory, when "belief" is understood in a less restricted sense.

The error theorist may reject premises (S₃) and (S₅) by rejecting (S₂) – the claim that the property of being a reason for belief is a normative property. She may do so by arguing that there are other senses of "reason for belief," e.g. hypothetical reasons. According to Streumer (2016b:423), this response invites the question why moral claims are not equivalent to descriptive claims. A related question is why, if epistemic norms can be subject to a Kornblith-style treatment (as suggested in chapter 5), moral norms cannot. But moral norms are *not* related to any goal that we have hypothetical reasons to care about as soon as we desire something. Moreover, our way of using and thinking of, moral claims fits poorly with the suggestion that they are equivalent to hypothetical imperative claims.

Even if the error theorist does not reject (S₂), she may attempt to reject (S₅) by rejecting (S₄) – the claim that we cannot fail to believe what we believe to be entailed by our beliefs. Olson has argued that the *preface paradox* constitutes a counterexample to (S₄). While it is not clear that this objection succeeds, it serves to highlight how restricted the unbelievability-claim is. As Olson points out, it is reasonable to suppose that error theorists in general do not believe error theory with the highest possible confidence, but merely believe it to some lower (although relatively high) degree. Because Streumer's argument is restricted to belief in error theory where such belief entails that the believer is

wholly (or very) confident, its success does not rule out such belief in error theory.

6.3 Unbelievability reconsidered

This last point made by Olson (2014:170) – i.e. that few defenders of error theory believe error theory with the highest possible confidence – highlights a final challenge I would like to raise to Streumer’s argument against the possibility of believing (normative) error theory.

We saw above that according to Streumer, an agent, *S*, can believe error theory only if *S* meets the following conditions:

- (1) *S* believes what *S* believes to be entailed by error theory,
- (2) *S* does not believe that there is no reason to believe error theory,

Error theory involves the claim that there are no normative properties. Consequently, if we grant Streumer that reasons for belief are irreducibly normative, and that *S* must meet conditions (1) and (2) to believe normative error theory, it follows that *S* cannot believe error theory.

But we have also seen above that (1) and (2) constitute plausible restrictions on belief in error theory only if understood as restricted to a specific form of belief in the theory. First, the belief in question has to be full as opposed to partial, non-compulsive as opposed to compulsive, and explicit as opposed to implicit. Second, the object or content of the belief in question must be the error theory itself, and not that error theory is true. Third, the believer must adequately understand the error theory. Conditions (1) and (2) do not apply to someone who believes error theory because, for example, a respected philosophy professor has told them that error theory is true.

Taking this into consideration, Streumer’s argument, if successful, does show that that it is impossible to believe error theory on a *natural but less restricted* interpretation of this claim. It merely shows that it is impossible to believe error theory on a *restricted* interpretation of this claim which implies that the belief is a full, non-compulsive, explicit belief (in the error theory itself), and that the believer adequately understands the error theory. The impossibility of being in such a state does *not* imply that it is impossible to hold a full, non-compulsive explicit belief *that error theory is true*, at least not for someone who does not adequately understand the error theory.

Given this, the conclusion of Streumer’s argument is not as radical as it may first appear. First and most obviously: When “belief” is understood in this restricted way, the truth of the unbelievability claim no longer seems to suggest an error theory with respect to thought and discourse about belief in normative error theory. It is possible for the error theorist to believe or claim that she

believes the error theory without believing or claiming that she holds a full, non-compulsive, explicit belief in the error theory itself (where such belief implies adequate understanding of the theory).

Second, I claimed above that the error theorist may argue that because we typically desire things, we typically have hypothetical reasons to follow epistemic norms, and therefore have hypothetical reasons to believe error theory. Moreover, she may use Olson's (2014) distinction between arguments to the effect that p is true, and arguments to the effect that there are reasons to believe p (see chapter 5) to argue that while the fact that there are no normative properties entails that there are no normative reasons of any kind to believe that error theory is true, this does *not* entail that error theory is not true, nor that that there cannot be sound arguments to the effect that error theory is true. With the help of Olson's distinction, the error theorist may offer arguments to the effect that error theory is true – i.e., arguments the premises of which, if true, guarantees or makes highly probable that error theory is true. In addition to this, she may argue that the premises in such arguments is *evidence* – in a non-normative sense of “evidence”, where something can be evidence for p without being a normative reason to believe that p – that error theory is true. The error theorist may even claim that some such argument is a *powerful* argument for the truth of error theory, because in analogy to how there are ways of understanding the notion of a good chess move according to which being such a move does not entail having a normative property, there are ways of understanding the notion of a powerful argument according to which being such an argument does not entail having a normative property (Olson 2014:160-3, 170 ff.). If the error theorist has formulated such an argument, it seems to me that she may have high confidence that error theory is true, where this confidence is based on evidence that makes it highly probable that error theory is true. If the error theorist can be in such a state – i.e., a state of high confidence based on evidence making it highly probable that normative error theory is true – it might not matter much to her that she cannot hold a full, non-compulsive, explicit belief in the error theory itself (where such belief implies adequate understanding of the theory).

6.4 Conclusion

This chapter has discussed what I have referred to as *the psychological objection* to ERROR THEORY. I have challenged the argument by indicating ways in which the normative error theorist may respond to Bart Streumer's argument for the claim that it is impossible to believe normative error theory.

I argued that the normative error theorist can reject premise (S₂) – and thereby (S₃) and (S₅) which depend on it – by understanding epistemic reasons along the lines suggested in chapter 5. She may argue that agents may have hypothetical reasons to believe error theory, and that when “reason to believe” is

understood along these lines, the property of being a reason for belief is not a normative property.

A second way in which the error theorist may respond to this argument is to take it on board and argue that this does not exclude the possibility of several modes of belief in normative error theory. The truth of the unbelievability claim does not exclude the possibility of being in a state of high confidence based on evidence making it highly probable that normative error theory is true. Given this, the unbelievability claim is not as radical as it may first seem. (In particular, it does not imply error theory with respect to belief in normative error theory.)

If these arguments are sound, there are ways for the error theorist both to reject the argument for the unbelievability claim and to take it on board. If so the psychological objection does not show that error theory (whether normative or moral) has entailments that cannot be believed.

Chapter 7

Against propagandism and conservatism

In the previous four chapters, I discussed two well-known arguments for ERROR THEORY – viz. what I have referred to as *the argument from non-institutionality* and *the argument from irreducibility* – and two objections to it – viz. what I have referred to as *the epistemic objection* and *the psychological objection*. I argued that neither of the arguments in favor of ERROR THEORY gives us strong reasons to believe it, and that neither of the objections to ERROR THEORY gives us strong reasons to reject it. Given this, it seems reasonable to suppose that the truth of ERROR THEORY remains a non-negligible possibility. The topic of this and the following chapter is the practical implications of ERROR THEORY for ordinary everyday moral thinking and discourse. More specifically, the central question of these two chapters is what we ought to do if we find out that ERROR THEORY (from now on, “error theory”) is true.¹

Different answers to this question have been suggested in the contemporary philosophical literature. Defenders of *abolitionism* suggest that we ought to abandon everything that resembles ordinary first-order moral thinking and discourse. Doing so would free moral practice of untrue moral beliefs and assertions, and thereby of error. Others suggest that we ought to continue with ordinary first-order moral practice as it is, or was, prior to the discovery that error theory is true (from now on, “(present) moral practice”), or something similar to it, because it is useful to do so.² A first such suggestion is

¹ Joyce (2005:287-8) and Olson (2014:178) pose similar questions. Köhler and Ridge (2013:3) claim that error theory about any domain of thought and discourse raises the question what to do with that domain.

² In the previous chapters, I have used the term “(ordinary) moral practice” to refer to that practice of first-order moral thought and discourse which we are familiar with from our everyday lives. In this chapter, I use “present moral practice” (or, context allowing, “moral practice”) to refer to this same practice *in the state it is in (or was in) prior to the discovery that error theory is true*. In contrast, I use “revised moral practice” to refer to moral practice in the state it is in *after being revised in the way fictionalism recommends*. (Because I am not assuming that error theory is true, but merely that its truth *remains a non-negligible possibility*, I understand the present state of moral practice as a state where it has not been discovered that error theory is true. My use of the phrase “present moral practice”, however, should not be understood as dependent on the assumption that it is not the case that error theory is true.)

propagandism, which says that if we find out that error theory is true, we ought to keep the truth of error theory secret, so that ordinary people may continue engaging in moral practice just like before, unaware of the error. A second is *conservationism*, which says that we ought to continue believing and asserting moral propositions in everyday contexts, and attend to our belief that error theory is true only in critical contexts such as that of the philosophy seminar room. A third is *fictionalism*, which says that we ought to carry on with morality as a fiction, and that doing so essentially involves revising present moral practice in such a way that it no longer involves any *genuine* moral beliefs or assertions but only moral *pretense-beliefs* and *pretense-assertions*.

If error theory is true, questions about what we *ought* to do might appear suspect. But as long as such questions do not concern what we *morally ought* to do, they need not be. (More on this in section 7.1, below.)

One of the aims of this and the following chapter is to argue that if moral practice is overall non-morally valuable, we ought to implement fictionalism – i.e., revise moral practice in the way fictionalism recommends – if we find out that error theory is true. Here are some initial reflections to make fictionalism attractive: First, as Simon Blackburn (1985:4) remarks, there is something fishy about on the one hand holding an error theory, yet, on the other hand continuing to moralize as usual. If moral practice embodies an error, it would be best if our old infected moral concepts and ways of thought were replaced by ones which simultaneously serve our needs and avoids the errors. Second, I argued in chapters 5, above, that we typically have hypothetical reasons to care about truth and truth-conduciveness and therefore have hypothetical reasons to evaluate doxastic states and actions in relation to these goals – i.e., that having true beliefs is typically instrumentally valuable to us. While this does not imply that untrue beliefs are necessarily instrumentally disvaluable to us, it at least suggests that if there are no true moral propositions, we typically lack instrumental reasons to believe such propositions. Third, moral practice seems instrumentally valuable to us in that it seems to play an important role in our lives. If this is the case, it is reasonable to suppose that we would care about preserving as much of it as possible – or at least that doing so would be in our interest – if we found out that error theory is true.

This chapter is organized as follows: Section 7.1 explains how the question of what we ought to do if we find out that error theory is true is understood in this and the following chapter. Section 7.2 argues that abolitionism comes at a cost, because present moral practice seems overall non-morally valuable. Section 7.3 argues that propagandism is incomplete, and has not provided sufficient reasons to prefer it over competing strategies whose implementation would leave present moral practice more or less as it is. Section 7.4 argues that conservationism has not provided sufficient reasons to prefer it over fictionalism.

The next chapter argues that a version of fictionalism which recommends that present moral practice is revised on the model of role-playing in *live action role-playing games* has potential to preserve benefits of present moral practice and handle some *prima facie* problems for fictionalism.

7.1 Preliminaries

This section explains how I understand the central question of this and the next chapter – i.e. the question what we ought to do if we find out that error theory is true.

A first thing to note is that we do not need to presuppose that error theory is *true* in order to discuss the question of what we ought to do if we find out that error theory is true. The only presupposition needed is that the truth of error theory *remains a non-negligible possibility* – as it surely does.

Second, if error theory is true, to suggest that we *morally ought* (e.g.) to implement fictionalism is to suggest something untrue. That is; if error theory is true, there can be a true positive answer to the question what we ought to do if we find out that error theory is true only if “ought” is understood as a *non-moral ought*. Because the discussion in this chapter presupposes that the truth of error theory is a non-negligible possibility, understanding the relevant “ought” as a *moral ought* is ruled out (for discussion, see Joyce 2001:177, 2005:288, Olson 2011b:184, 2014:183 and Hussain 2010:340-1).

In the context of discussing what we ought to do if error theory is true, I use “ought” to express what best satisfies our shared preferences. In doing so, I do not make any commitment about the semantics of our common practical uses of “ought” (i.e., I am not assuming that this is a common practical use). I believe that the question of what course of action best satisfies our shared preferences if we find out that error theory is true is interesting independently of the semantics of our ordinary non-moral uses of “ought”.³ Consequently, in arguing that

we *ought* to implement fictionalism if we find out that error theory is true,

I am defending the claim that

implementing fictionalism is the course of action that *best satisfies our shared preferences* if we find out that error theory is true.

³ Joyce says that the question what we should do if moral error theory is true invites an answer in the form of a hypothetical imperative, and that the relevant “ought” is “just a straightforward, common-or-garden, *practical* ‘ought’ (Joyce 2005:288).” For discussion of this claim, see Olson (2014:183).

In the next chapter, I provide considerations given which implementing fictionalism is the course of action that *best* – i.e., most efficiently and to the greatest extent – contributes to the satisfaction of our shared preferences if we find out that error theory is true. In doing so, I do not assume that these considerations provide *objectively authoritative* or *normatively irreducible* reasons for accepting fictionalism (see chapters 3 and 4).

It can be argued that this way of approaching the question of what practical implications the discovery that error theory is true has for ordinary everyday moral thinking and discourse is implicit also in the suggestions made by Mackie in chapter 5 of *Ethics*. Mackie argues that although we have discovered that error theory is true, the human predicament is still such that we need a device for counteracting our limited sympathies. However, this set of constraints on our conduct – this *morality*

[...] is not to be discovered but to be made: we have to decide what moral views to adopt, what moral stands to take. [...] [The object of this exercise is to] decide what to do, what to support and what to condemn, what principles of conduct to accept and foster as guiding or controlling our own choices and perhaps those of other people as well (Mackie 1977:106).

Mackie identifies a preference shared among humans – viz., that of counteracting limited sympathies – and suggests that what strategy to implement – i.e., what moral views to adopt, what principles of conduct to accept and foster as guiding or controlling our own choices and perhaps those of other people as well – is to be determined by what strategy best satisfies this preference; the strategy implemented shall be a device for satisfying this preference.

Third, because the central question in this chapter is what strategy *best* satisfies our shared preferences if error theory is true, it is possible that the best strategy does not succeed in preserving *all* benefits of present moral practice (Joyce 2005:299, 302).

Fourth, I follow Joyce (2001:177, 2005:288) in understanding the question what *we* ought to do if we find out that error theory is true as the question what *a group of people who share some desires, enterprises, needs, aims, etc., ought to do* if they find out that error theory is true (where “ought” is understood in the way suggested above).

7.2 Abolitionism and the value of moral practice

This section argues that abolitionism comes at a cost, because moral practice seems to be overall (non-morally) valuable in that it provides some pretty serious benefits throughout ordinary life.

Abolitionism (or “*eliminativism*”) says that if we find out that error theory is true, we ought to abolish everything that resembles moral thought and discourse: moral reasoning, thinking, judging, the making of moral statements, etc. (Garner 2010:224, 226, 232-3).⁴ Abolitionism seems like a natural response to the discovery that error theory is true: if moral practice is altogether abolished, we can be sure that its errors are avoided.

Fictionalists, propagandists and conservationists believe that we are, despite the truth of error theory, better off with moral practice in some form. Abolitionists, in contrast, believe that we are better off without it. Garner suggests that any benefits gained by having a moral practice are outweighed “by the harm that comes from having to promote and defend a series of easily questioned falsehoods (Garner 2010:224),” and that although even the “abolitionist may find an occasional moralistic utterance overwhelmingly useful (ibid:222),” he is “more impressed by the harm of not seeing the error as an error (ibid).” We will see below that fictionalists and propagandists agree with abolitionists that it is bad to hold untrue beliefs and to make untrue assertions (conservationism, in contrast, recommends it). But they do not agree that this motivates a complete abolishment of everything that resembles moral thinking and discourse.

It may be argued, in favor of abolitionism, that moral thought and discourse is not necessary for ending things such as genocide, slavery, cruelty, etc. Garner suggests that to eliminate genocide and slavery, the essential thing is not that we are able to say that genocide and slavery is morally wrong, but that we *feel* a deep and sincere aversion towards it. Agents may hold such aversion without holding moral beliefs (Garner 1993:98). Fictionalists will probably agree that moral beliefs are not necessary for ending genocide (etc.). As Joyce (2013:2) points out, for someone who strongly *desires* to live in a world without genocide, it is not obvious that the force of her desire should weaken if she were to stop categorizing genocide specifically in moral terms.⁵ Moreover: For someone who is moved by the suffering, death, and extinction of vast swaths of human organization and culture, it might be that the thought that she is morally required to do something about it does little beyond this. It at least seems possible to have urgent sense that something must be done without thinking that this is morally required. What would happen if moral thoughts and beliefs were universally abolished is of course an empirical question. But given these considerations, it at least seems possible that motivation remains without moral thoughts and beliefs.

⁴ The discussion below is limited to what Olson (2014:179 ff.) refers to as “complete” abolitionism. Complete abolitionism is the recommendation, of e.g. Ian Hinckfuss (1987) and Richard Garner (1993), (2010), that moral practice is dropped in its entirety. Partial abolitionism is the recommendation that some subset of moral practice is dropped and/or replaced.

⁵ See also Joyce (2007:57-8).

The value of moral practice

The main problem with abolitionism is that even if it is possible for motivation to end genocide (etc.) to remain without moral thought and discourse, abolishing morality seems to come at a pretty serious cost. Or so it has been argued. In the literature covering the issue of what we ought to do if we find out that error theory is true, almost everyone seems to accept that (present) moral practice has some functions or features whose preservation is desirable. The value of these functions or features is usually taken to consist in their being instrumentally or practically useful, or in their serving some (non-morally) valuable purpose. Several writers suggest that moral practice is, in this way, *valuable as a means* for things we value, desire or care about.⁶

Moral practice seems valuable *for social and societal human relations*. Mackie, for example, suggests that

We need morality to regulate interpersonal relations, to control some of the ways in which people behave towards one another (Mackie 1977:43).

Even if we do not agree with Mackie that the human situation is fundamentally one of partial conflict (Mackie 1980:156), and that people are almost always more concerned with their selfish ends than with helping others (Mackie 1977:107), it seems clear that this is at least sometimes the case, and that morality is a device among others that can help resolve or mitigate partial conflicts (Mackie 1980:156), and help counteracting limited sympathies (Mackie 1977:108). Daniel Nolan, Greg Restall and Caroline West similarly claim that “[m]orality plays an important social role in coordinating attitudes and regulating interpersonal relations (Nolan, Restall and West 2005:307).” It has also been suggested that moral thought and discourse is central for solving collective decision problems in mutually beneficial ways, by help solving coordination problems and familiar types of prisoners’ dilemma situations (Mackie 1985:157 ff.; Köhler and Ridge 2013:6).

Moral practice seems practically valuable also *at an individual level*. Köhler and Ridge suggest that normative thought and discourse *in general* is important for reflective and coherent planning in the face of temptation, and individual decision making (Köhler and Ridge 2013:6-7). For example, normative thoughts helps us structure our practical deliberation “by putting certain decisions off the table (ibid:6).” If this is correct, it is reasonable to suppose that the same kinds of benefits are associated also with normative thought and discourse specifically of the moral kind.⁷ Joyce suggests that

⁶ Abolitionists may agree with this claim while denying that it justifies *preserving* moral practice (Campbell 2014:448).

⁷ I am not denying that the set of decisions put off the table by moral thoughts might differ from that put off the table by non-moral normative thoughts.

morality helps us act instrumentally rational by helping us escape weakness of will, passion, negligence, etc., and other things that obstruct the achievement of our long-term desires (Joyce 2001:181, 184, 210 ff., 2005:300-2).⁸ According to him, moral beliefs help us act instrumentally rational because

[i]f a person believes ϕ ing to be required by an authority from which she cannot escape, if she imbues it with a “must-be-doneness,” if she believes that in not ϕ ing she will not merely frustrate herself, but will become reprehensible and deserving of disapprobation – then she is more likely to perform the action (Joyce 2001:184).

In short: moral thought and discourse is valuable for us in virtue of the practical roles it plays, and practical work it does, for us. It is valuable for social and societal human relations, for helping us structure our practical deliberation, and for attaining things we value, desire or care about.

Some abolitionists suggest that moral practice should be abandoned on other grounds than those provided by the (potential) truth of error theory, and that abolishing morality would be a good thing *overall*. It has been suggested that morality can be used for objectionable purposes such as defending the sanctions of tyrants, that it inflames disputes and makes compromise difficult, and makes global war possible (Garner 2010:219-221).⁹ It has also been suggested that the more the members of a society are motivated by moral considerations, the more likely is it that this society is elitist, authoritarian, intellectually dishonest in its social decision making, etc. (Hinckfuss 1987).

While I do believe that morality is overall beneficial, and that these challenges are worth taking seriously, I cannot begin to seriously address the empirical issue of the overall benefits or dangers of morality here. A serious discussion of these issues would simply lead too far. For the sake of the discussion in this and the following chapter, I will proceed under the assumption that present moral practice is overall valuable in that it provides pretty serious benefits throughout ordinary life, as indicated above.

7.3 Against propagandism

This section argues that propagandists have not yet provided a complete response to the question what we ought to do if we find out that error theory is true, nor sufficient reasons to prefer propagandism over competing strategies

⁸ For discussion of this claim, see Miller (2013:115 ff.).

⁹ Garner refers to Mackie’s (1980) claim that although morality fulfils a social function, it has some negative side-effects. What Mackie ultimately suggests, however, is that morality is one device among others that can help resolve social conflicts (see the quote above, and Mackie 1980:155-6).

whose implementation would leave moral practice (more or less) as it was prior to the discovery that error theory is true.

Propagandism says that if we find out that error theory is true, we ought to keep this fact a secret and “hush up” any evidence for it, so that ordinary people may carry on engaging in moral practice just like before, unaware of its error (Joyce 2001:214; Cuneo and Christy 2011:93). Propagandism presupposes that if it is discovered that error theory is true, this will not be a collective discovery across a society. Rather, this discovery will be made by a limited group of sufficiently (philosophically) sophisticated people (Cuneo and Christy 2011:94).¹⁰ For ease of presentation, I will from now on follow Terence Cuneo and Sean Christy in referring to this group as “the elites”, and to the group of people who are not elites as “the folk” (see e.g. Cuneo and Christy 2011:93).

Fictionalists, propagandists, and conservationists all agree that moral practice has benefits that are desirable to preserve if error theory is true, but disagree on what *means* we ought to take to preserve these benefits if we find out that error theory is true. While fictionalists believe that the best strategy is to *revise* moral practice in a way that both remedies the error and preserves (at least some of) its benefits, propagandists and conservationists believe that it is best to *leave moral practice more or less as it was prior to the discovery that error theory is true*.¹¹ While this strategy does not remedy the errors of moral practice, it seems like a straightforward way of preserving its benefits: if moral practice is left (more or less) as it is, there seems to be no risk of losing any of its benefits. However, both propagandism and conservationism encounter problems that concern their practical effectiveness.

Three objections to propagandism

It has been argued that propagandism is practically unstable. Köhler and Ridge suggest that it is likely that in the long run, the truth will come out:

We assume here [...] that the arguments for [...] [error theory] have become sufficiently clear and powerful that few would deny their force if exposed to them. This reflects that we are after a *contingency plan*, which would only be appropriately implemented if we became rightly very confident that [...] [error theory] is true. This, though, presumably would only happen if it became clear that the arguments for the error theory were very strong. In this scenario [...], there is something farcical

¹⁰ The central question of the present chapter is what *we* ought to do if we find out that error theory is true, where “we” is understood as a *group of people who share some desires, enterprises, needs, aims, etc.* Propagandism can be understood as imposing a restriction on the interpretation of the relevant “we”.

¹¹ This similarity between propagandism and conservationism is also pointed out by Olson (2014:196).

about the idea of a small elite keeping the truth hidden (Köhler and Ridge 2013:11).¹²

Joyce similarly claims that no policy that encourages the promulgation of false beliefs in others will be stable in the long run. Moreover, implementing propagandism is too risky. If the truth would come out, “the result would be a very confused group of people, unsure of what to believe, and unable to trust their normal belief-producing mechanisms (Joyce 2001:214).”¹³

Aware of this critique, Cuneo and Christy seek to improve propagandism by adding the following two claims to it:¹⁴

- (1) [...] while those aware of the error should be prepared to cover up evidence of it were the need to arise, the need in fact rarely arises (Cuneo and Christy 2011:94).
- (2) [...] those aware of the error must often engage in moral deliberation with the folk. Propagandists [...] recommend that the elite take up various types of non-doxastic stances toward moral propositions, employing them in ordinary speech [...] to do such things as encourage, edify, or blame their interlocutors. Admittedly, to the initiated, the linguistic expressions of these non-doxastic stances will often appear to be one or another species of assertion. But in reality they are not. To use the vernacular of speech act theory, when engaging in ordinary moral discourse, the elite present moral propositions not so as to assert them (or pretend to assert them) but to have various kinds of perlocutionary effects on their audience [...] (Cuneo and Christy 2011:94-5).

Let us start by considering (1). Cuneo and Christy defend this claim by suggesting that the folk lack *the requisite ability and mindset* to understand error theory and the arguments for it:

¹² Although Köhler and Ridge (2013) discuss propagandism as a potential response to the question what is to be done if *normative* error theory turns out to be true, their criticism applies to propagandism understood as a potential response to the question what we ought to do if we find out that *moral* error theory is true.

¹³ Joyce is quoting Garner (1993:96) here.

¹⁴ Köhler and Ridge (2013:10-11) argue that an open defense of propagandism counteracts its chances of success, and that this raises the question why Terence Cuneo and Sean Christy are vocal about their adherence to the view. While it does seem reasonable to question someone’s openness about their adherence to propagandism if they are presently practicing it, I understand Cuneo and Christy as offering propagandism merely as a response to the question what we ought to do *if* we find out that error theory is true. Because I do not understand them as presupposing that error theory is true, I do not understand them as *presently practicing* propagandism.

[...] [T]o appreciate the nature of the error and the reasons for believing that it has been committed requires [...] a level of conceptual sophistication not possessed by most people. [...] [Moreover,] appreciating the error requires being open to seriously considering of views that are opposed to deeply entrenched convictions about morality – convictions that are often grounded in religious belief and practices. Many of the folk do not exhibit openness of this sort. If propagandists [...] are right about all this, attempting to communicate to the folk what is at stake in metaethical debates and what should be done about it is not worth the trouble (Cuneo and Christy 2011:94).

My first objection to propagandism is that this description of the folk risks making the propaganda-part of propagandism otiose.¹⁵ If the folk are *unable to understand error theory and the arguments for it*, there seems to be no need for a “hush-up” – even if the folk were presented with error theory and the arguments for it, they would not understand them, and would therefore continue believing and asserting moral propositions like before. Similarly, if the folk are *not open to seriously considering views that dispute their current convictions about morality*, there seems to be no need for the a “hush-up” – even if the folk were able to understand both error theory and the arguments for it, they would not be open to accept that error theory is true, and would therefore continue believing and asserting moral propositions like before. Cuneo and Christy’s description of the folk risks making the propaganda-part of propagandism otiose because if it is correct, moral practice will be left (more or less) as it was prior to the discovery that error theory is true, *regardless of whether the elite takes steps to hush up the fact that error theory is true and the evidence for it*.

Cuneo and Christy may be right that there is a degree of difficulty in understanding error theory and its support that, contrary to what some of the above criticism seems to presuppose, makes it relatively easy to keep it a secret. But perhaps some of the folk *would* understand the theory and the arguments for it if presented with them. And perhaps the amount of people who would do so is sufficiently high for it to matter how they would react to finding out that error theory is true. Perhaps this is enough to call for some constraint from philosophers who have seen the light.

Let us now turn to (2). Unfortunately, the passage quoted is not entirely simple to comprehend. Cuneo and Christy seem to make two recommendations concerning how members of the elite is to behave when engaging in moral practice with the folk after the elites have found out that error theory is true. First: when engaging in moral *deliberation* with members of the folk, the elites does not believe moral propositions. But in contrast to what fictionalism

¹⁵ Another objection that could be raised to this description of the folk is that it is overly pessimistic.

recommends, they does not pretend to believe them either. Instead, the elites have various other types of non-doxastic stances towards moral propositions. Second: when engaging in moral *discourse* with the folk, the elites makes moral utterances without making moral assertions, and with the intention to have certain perlocutionary effects on her audience.

It is in particular the second recommendation that is difficult to comprehend. If the elites make moral utterances while intending to be taken as making assertions the acceptance of which involves making a moral judgment corresponding to the content of the terms involved in the assertion, based on the assumption that their utterances are sincere, it seems hard to deny that they are making (insincere, defective) assertions.

Unfortunately, Cuneo and Christy do not elaborate on the issue of what a member of the elite is doing when she makes moral utterances in the presence of the folk. Perhaps she performs a speech act of some other kind than assertion. Or perhaps, given that Cuneo and Christy accept the idea that if a speaker makes a certain kind of utterance in the relevant context, the utterance is an assertion whether or not she intends it to be one, members of the elite make moral utterances without intending to make non-defective assertions or to make non-defective speech acts of any other kind (where an act is a defective speech-act if it satisfies conditions for being such an act, but does not satisfy constitutive norms for that act type).¹⁶ These details will not matter for the discussion below. The important point is that propagandism recommends that when a member of the elite engages in moral discourse with members of the folk, she makes moral utterances without making assertions, and with the intention of having certain perlocutionary effects on her audience.

My second objection to propagandism is that it is not sufficiently motivated. Cuneo and Christy suggest that

[a]ll else being equal, having massive amounts of false beliefs is bad. (We assume, for the moment, that the value of truth is not merely instrumental. There is, we assume, something non-instrumentally worthwhile about getting into cognitive contact with reality.) Accordingly, if someone were to find himself with packs of false beliefs about what Locke called ‘matters o maximal concernment,’ such as morality and religion, then he should want to remedy this (Cuneo and Christy 2011:95).

While the claim that *untrue beliefs are bad* supports propagandism’s recommendation that the elites do not believe moral propositions (whether they are engaged in moral deliberation with the folk or amongst themselves), it clearly lends no support to propagandism’s further recommendation that the

¹⁶ Thanks to Gunnar Björnsson for suggesting this interpretation.

elites continue making moral utterances when engaging in moral discourse with the folk. Presumably, the latter recommendation is motivated by the idea that doing so is a means of *preserving benefits of present moral discourse*. But although this aim seems to speak in favor of a general recommendation to continue using moral language, it lends no support to propagandism's further, more specific recommendation that the elite do so *without making assertions*. This is a problem for propagandism because there are other options as to how the elites may continue using moral language without believing any moral propositions.

One such option is lying – i.e., making *genuine but insincere* assertions.¹⁷ Suppose that Jessica is a member of the elite who is engaged in moral discourse with members of the folk. Suppose that Jessica lies when making moral utterances in their presence. Being members of the folk, her audience knows neither that error theory is true, nor that Jessica knows that error theory is true, nor that Jessica does not have any moral beliefs, etc. Given this, it is reasonable to suppose that they will often interpret Jessica's moral utterances as assertions. Suppose now that instead of lying, Jessica behaves in the way propagandism recommends, when using moral language in her audience's presence. If so, her behavior is at least misleading, for as stated in (2), her audience will often interpret her utterances as assertions if she behaves in this way. It is possible that it is *morally worse* to lie than to mislead. But if error theory is true, this is beside the point. The important point is that *regardless* of whether Jessica behaves in the way propagandism recommends or lies when engaging in moral discourse with the folk, they will interpret her utterances as assertions. But if they interpret her utterances as assertions, it is reasonable to suppose that her behavior will induce genuine moral beliefs and thoughts in them. Given this, it seems that regardless of which strategy Jessica chooses, her behavior will serve the aim of preserving the benefits of present moral discourse. Consequently, it seems that the aim of preserving these benefits does not give us reason to prefer propagandism's recommendation on how members of the elite are to behave when engaging in moral discourse with the folk to the competing recommendation that they lie.

It seems to be open to propagandists to respond to this objection by taking this on board – i.e., by letting go of the recommendation that the elites do not make moral assertions when engaging in moral discourse with the folk, thereby allowing the elites to lie in such contexts. Alternatively, propagandists could opt for a mixed strategy, for example one which combines propagandism and a fictionalism of the kind that will be suggested (in chapter 8) below. According to this suggestion, members of the elite would role-play themselves as moralizers when engaging in first-order moral practice with members of the

¹⁷ It is in other words supposed here that lying is a case of “abuse” rather than “misfire”. For more on the distinction, see Green (2017:§2.2). For a defense of the claim that lying is a species of assertion, see Joyce (2001:202).

folk. The propaganda-part would consist in the fact that the elite did not attempt to convince the folk that error theory is true. If the question of the truth of error theory came up, the elites would not attempt to convince the folk of the truth of error theory. Instead, they would attempt to convince them of the benefits of fictionalism. Note that given this suggestion, it is still possible that the moral utterances – i.e., *pretense-assertions* (see below) – made by the elite in contexts where fictionalism is not part of the common ground would end up being (insincere) assertions.

My third and final objection to propagandism is that it is not a complete theory. Propagandism does not provide a complete response to the question what we ought to do if we find out that error theory is true, for it gives no account of how members of the elite are to behave when engaging in moral discourse *amongst themselves*. As transpires from the above discussion, propagandism focuses on issues that concern how the elites are to behave to make sure that the folk continue enjoying the benefits of moral practice, but provides no account of how the elites are to behave to ensure that *they themselves* continue doing so. One implication of this is that propagandism does not sufficiently take into consideration the instrumental value of moral practice. If moral thought and discourse plays important practical roles and does important practical work for us, it is reasonable to expect that we would want to get hold of its effects also in contexts where the truth of error theory is accepted. But propagandism gives no account of how this is to be achieved. This makes it necessary to consider alternatives to it.

In this section, I have argued that propagandism does not provided a complete response to the question what we ought to do if we find out that error theory is true, and has not yet provided us with sufficient reasons for preferring propagandism to competing strategies whose implementation would leave present moral practice (more or less) as it is. One such competing strategy is conservationism, to which I will now turn.

7.4 Against conservationism

This section argues that conservationism has not yet given us sufficient reasons to prefer it over fictionalism, since it has not yet provided us with sufficient reasons to think that implementing it is better as a means of preserving benefits of moral practice than implementing fictionalism.

Conservationism, which is defended by Olson (2011b, 2014), says that if we find out that error theory is true, we ought to continue engaging in ordinary moral thinking and discourse just like before – i.e., continue holding moral beliefs and continue making moral assertions – and attend to our belief that error theory is true only in *critical contexts* (Olson 2011b:193, 199, 2014:178, 184, 192). Critical contexts are those where we engage in critical thinking and

serious contemplation of metaethical issues.¹⁸ I refer to any context that is *not* a critical context – including contexts of present or revised moral practice – as an “uncritical context.”

I claimed above that in recommending that we leave present moral practice (more or less) as it is, both conservationism and propagandism seem to offer straightforward ways of preserving its benefits. But unlike propagandism, conservationism provides a complete account of what we ought to do if we find out that error theory is true.

Let us start by considering conservationism’s recommendation that we continue holding genuine untrue moral beliefs if we find out that error theory is true. Olson says that

[...] conservationism recommends moral belief in morally engaged and everyday contexts and reserves attendance to the belief that moral error theory is true to detached and critical contexts, such as the philosophy seminar room (Olson 2011*b*:199, 2014:192).

Given this, the recommendation is *not* that we believe error theory only in critical contexts, and continue believing (say) non-naturalist moral realism in (uncritical) everyday contexts. Rather, it is that we *simultaneously* believe *both* (e.g.) that it is morally wrong to torture animals, and that because there are no moral properties, nothing is morally wrong, but that we *attend* to the former belief only in uncritical contexts, and to the latter only in critical contexts (Olson 2011*b*:199-200).

Consider, for illustration, the proposition that *it is (morally) wrong to torture animals*. Call this proposition “*p*”. Suppose that error theory’s account of present moral practice is correct, and that Mark is like most other people. Given this, Mark both believes and is willing to assert that *p* when engaging in ordinary moral thinking and discourse. Suppose that when taking a class in metaethics, Mark finds out that error theory is true. Because *p* is a moral proposition, he starts believing (at least implicitly) that *not-p*. Conservationism recommends that despite acquiring this new belief, Mark holds on to his belief that *p*.

Although conservationism recommends Mark to simultaneously believe both that *p* and that *not-p*, it does not recommend him to hold two inconsistent *occurrent* beliefs. What it recommends is rather that after finding out that error theory is true, Mark has an *occurrent* belief that *p* and a *dispositional* belief that *not-p* (i.e., has a disposition to believe that *not-p* in critical contexts) when

¹⁸ The example of critical contexts most frequently referred to in the literature is *the philosophy seminar room*. Other examples include the court room, and various pedagogical contexts. See Joyce (2001:190 ff., 2005:289-90), Olson (2011*b*:199, 2014:192), Cuneo and Christy (2011:87-90) and Köhler and Ridge (2013:15).

engaging in ordinary moral practice.¹⁹ But even given this, one might wonder whether it is possible for someone to “be convinced – in the seminar room – that nothing is morally wrong [...] and also to be convinced – out of the seminar room – that some acts are morally wrong (Olson 2011b:199).” In defense of the possibility of this, Olson claims that it is a psychologically familiar fact that we sometimes temporarily believe things we are disposed to disbelieve in critical contexts. For example,

[...] someone might say truly the following about a cunning politician: ‘I knew she was lying, but hearing her speech and the audience’s reactions, I really believed what she said’. Or a deceived lover might say about his mistress, ‘I knew she was lying, but when she told me that she cared about me I really believed her’. Hence we are sometimes *taken in* by what people say (be it cunning politicians, manipulative partners, etc.) in the sense that we believe what is said, even though we are disposed to believe, upon detached and critical reflection, that it is false (Olson 2014:192-3).

Olson suggests that something similar might be going on with moral beliefs. In analogy to how *other people* may seduce us into believing things they say although we are disposed to believe (in critical contexts) that they lie, *our own feelings* may seduce us into believing that things have moral properties although we are disposed to believe (in critical contexts) that error theory is true:

The error theorist might say, ‘I knew all along there is no such thing as moral wrongness, but hearing on the news about the massacre on civilians, I really believed that what the perpetrators did was wrong; I really believed that the UN ought morally to enforce a cease fire’. Or at a more personal level, ‘I knew all along that there is no such thing as moral requirements, but when I realized that breaking the promise would badly hurt his feelings I came to believe I was morally required not to break it’ (Olson 2014:193).

An objection to conservatism

The main problem with conservatism is that it has not yet given us sufficient reasons to prefer it over fictionalism, since it has not yet provided us with sufficient reasons to think that implementing it is better as a means of preserving benefits of present moral practice than implementing fictionalism.

¹⁹ For more on the distinction between occurrent and dispositional belief, see Schwitzgebel (2015).

There are several similarities between conservatism and fictionalism. First and most obviously, conservatism and fictionalism are competing suggestions as to what means we ought to take to preserve benefits of present moral practice if we find out that error theory is true. So on both views, present moral practice has benefits that are desirable to preserve if we find out error theory is true. Second, both views presuppose that a distinction can be made between critical and uncritical contexts. Third, they recommend similar behavior. For critical contexts, their recommendations seem identical: both views recommend that we attend to our belief that error theory is true, and do not make any moral assertions. There is also some overlap in their recommendations for uncritical contexts: both views recommend us to think moralized thoughts (i.e., think or say to ourselves e.g. that it is wrong to torture animals) and to use moral language while holding a dispositional belief that nothing is morally wrong. Fourth and last, both views recommend what Olson refers to as a “moral compartmentalization” or “two-level thinking”, since they recommend that we reserve attendance to our belief that error theory is true to critical contexts (Joyce 2001:190-4, 2005:289-90; Olson 2011b:199, 2014:192).²⁰

The essential difference between conservatism and fictionalism concerns what attitude to moral propositions and what use of moral sentences they recommend for uncritical contexts. While both views recommend that we think moralized thoughts and make moral utterances when we engage in ordinary moral practice after finding out that error theory is true, conservatism recommends that we continue to *genuinely* believe and assert moral propositions in such contexts, while fictionalism recommends that we change our attitudes to moral propositions and our way of using moral sentences so that we no longer genuinely believe or assert moral propositions, but merely *pretend* to do so. (More on this in chapter 8.)

We saw above that Olson defends the possibility of simultaneously holding both an occurrent moral belief and a dispositional belief that error theory is true by appealing to the familiarity of examples of scenarios where

- (i) in an uncritical context, the agent, *S*, says or thinks to herself that *p* while holding a dispositional belief that *not-p*,
- (ii) in a critical context, *S* reflects on her previous behavior and reports that she *believed that p* in the uncritical context.

But there are also scenarios where

²⁰ At least where “fictionalism” picks out the kind of fictionalism defended by Joyce. In chapter 8, I defend a version of fictionalism according to which no such thing is needed.

- (i) in an uncritical context, *S* says or thinks to herself that *p* while holding a dispositional belief that *not-p*,
- (ii*) in a critical context, *S* reflects on her previous behavior and reports that she believed that *not-p* in the uncritical context.

Examples of scenarios of the latter kind are congenial to Joyce's claim that as long as the agent is disposed to believe that *not-p* in *critical contexts*, she believes that *not-p* also in *uncritical contexts*, regardless of what she says or thinks to herself in the latter kinds of contexts (Joyce 2001:192-3, 2005:289-90). Moreover, it seems to me that such scenarios are at least as familiar as those appealed to by Olson.

Suppose, for illustration, that Tony is shopping for canned soup. He has never tasted canned soup before, but finds *Heinz* ketchup savory. Standing in front of the shelf, Tony says to himself "if their ketchup is good, their soup is sure to be good too", and puts a can of *Heinz* mushroom soup in his basket. Suppose that in this moment, someone approaches Tony and asks, in a serious tone of voice, whether he really believes that it is correct to generalize from a single experience with a savory *Heinz* product that all *Heinz* products are savory. Tony pauses to think for a moment before replying that he has *never* believed this, but that he chooses to take this for granted when shopping because it makes shopping easier and more convenient.

Given that Tony can take for granted things he does not believe when shopping, it seems plausible to suppose that the error theorist can take for granted things she does not believe when engaging in ordinary moral practice (perhaps because it makes practical decision making, solving collective decision problems, etc. easier and more convenient). If she can do this, she can think moralized thoughts when engaging in ordinary moral practice without holding any genuine moral beliefs.

Olson might respond that what Tony realizes in the critical context is not that he *never believed* that it is correct to generalize from single observations, but rather that something he *really believes* (or *did really believe*) in uncritical contexts is not true.²¹ Olson may be right that Tony genuinely believes the relevant proposition, *p*, in such contexts. But it at least seems *possible* to me that Tony's self-report is correct. If so, he *does not believe* that *p*, but merely thinks or says to himself that *p*, in uncritical contexts (where saying or thinking to oneself that *p* does not entail that one believes that *p*.)

I would like to make two points in relation to this. First, it seems that we cannot draw any general conclusions about what attitudes an agent takes to moral propositions in uncritical contexts simply from the fact that she thinks moralized thoughts and holds a dispositional belief that error theory is true in

²¹ Olson pointed this out in a comment to an earlier version of this chapter.

such contexts. It might be that she holds genuine moral beliefs, like Olson suggests. But it might also be that she does *not* hold any genuine moral beliefs, like Joyce suggests. Second, and perhaps more importantly, moralized thoughts seem to have some bearing on motivation and action *regardless of whether the agent holds any genuine moral beliefs*.²² It seems probable that the agent who is seduced by the rhetorically skilled politician is at least ready to affirm – when still in the uncritical context – that he believes what the politician says. It also seems probable that at least part of the reason why Tony puts the can of *Heinz* mushroom soup in his basket is that he is saying to himself that if their ketchup is good, their soup is sure to be good too. This suggests that regardless of whether we adopt conservatism or fictionalism, our moralized thoughts will affect our motivation and behavior *at least to some extent*.

Olson offers two reasons for preferring conservatism's strategy for preserving benefits of present moral practice over fictionalism's. The first is that it is not clear that engagement with the moral fiction can have bearing on action and motivation without causing genuine moral beliefs. The second is that if the resolution to adopt the moral point of view is a *precommitment*, like Joyce suggests that it is, the point of engaging with the moral fiction becomes unclear, and fictionalism's practical recommendations seem to pull in opposite directions.

In chapter 8, I spell out and defend a version of fictionalism which has potential to handle these objections. This fictionalism is similar to Joyce's but differs from it in suggesting that present moral practice is to be revised in such a way that engagement in revised moral practice is analogous to role-playing in a live action role-playing game (agents participating in revised moral practice *role-play themselves as moralizers*). This leaves us (presently) with the task of determining whether conservatism or fictionalism has the best potential of preserving benefits of present moral practice.

In relation to this, I would like to make two points. First, conservatism is preferable to fictionalism only if implementing conservatism is a better means of preserving benefits of present moral practice than implementing fictionalism. Second, if present moral practice is non-morally valuable and fictionalism is preferable to conservatism, then fictionalism is preferable to abolitionism and propagandism as well.²³

²² I take Olson to acknowledge this when he argues that we ought to go on thinking moralized thoughts after finding out that error theory is true *even if* we were unable to continue holding genuine (first-order) moral beliefs. According to him, thinking or saying to oneself (e.g.) that stealing is morally wrong might bolster self-control in the face of temptation *even if* one believes this to be false (Olson 2014:195).

²³ To see this, consider the current dialectic: We are trying to figure out what we ought to do if we find out that error theory is true. There are four competing answers to this question, represented by four competing theories: abolitionism, propagandism, conservatism and fictionalism. I started by arguing that abolitionism comes at a cost, because moral practice seems to be overall non-morally valuable in that it provides some pretty serious benefits throughout ordinary life. So abolitionism was (tentatively) ruled out. This left us with three competing suggestions as to how the benefits of

Because it seems to me that conservationism cannot preserve moral motivation to the same extent as fictionalism of the kind I prefer, it seems to me that conservationism is not preferable to fictionalism.

Olson worries that fictionalism will not preserve the benefits of present moral practice because it is not clear whether engagement with the moral fiction can have a bearing on motivation and action without causing genuine moral beliefs, and because the error theorist engaged in revised moral practice will sometimes have to attend to her belief that morality is a fiction. I will address these worries later on. But it seems to me that conservationism faces a similar problem.

According to Olson, it is possible for error theorists to simultaneously hold occurrent moral beliefs and dispositional beliefs that error theory is true, because

[j]ust as people might be seductive to the effect that it seems to us that what they say is true and that we virtually cannot help believing what they say, certain actions and events may engage our emotions of anger, empathy, etc., to the effect that it seems to us that the actions are morally wrong and that we virtually cannot help believing that they are morally wrong, no matter how intellectually compelling we find arguments in favour of moral error theory. It appears realistic that in morally engaged and engaging contexts, affective attitudes like anger, admiration, empathy, and the like, tend to silence beliefs that moral error theory is true (Olson 2014:193).

But even if this is true, there is reason to doubt that implementing conservationism is a better means of preserving benefits of present moral practice than implementing fictionalism, for there is reason to doubt that conservationism provides *robust* moral motivation. A common feature of Olson's examples is that they describe situations where an agent is under strong pressure to believe some proposition – so strong that she *virtually cannot help* believing it despite being disposed to disbelieve it critical contexts. The agent is *taken in* or *seduced* – by a rhetorically skilled politician and the reactions of the audience, by a manipulative partner, or by her own emotional reactions to a news-report about utterly horrible crimes against humanity or to the realization that acting in some way would badly hurt the feelings of someone she cares about – into *temporarily believing* something. If the error theorist's being

moral practice are to be preserved. I argued that propagandism does not (yet) provide a complete response to the relevant question, and that there seems to be no reason to prefer propagandism to competing strategies whose implementation would leave present moral practice more or less as it was prior to the discovery that error theory is true, like conservationism. With propagandism ruled out, we are left with two competing suggestions as to how the benefits of moral practice are to be preserved if we find out that error theory is true: conservationism and fictionalism.

motivated to act according to moral norms depends on her holding an occurrent genuine moral belief, and her being in such a state depends upon her being temporarily seduced by her emotional responses, moral motivation seems to hang on a thread. It is one thing to claim that when under strong pressure, the error theorist is sometimes temporarily seduced into believing that some action or event has some moral property, another to claim that she is *repeatedly* and *reliably* so seduced in contexts of ordinary thought and discourse. While I am ready to agree with Olson on the truth of the first of these claims, I have some reservations against the second.

First, even given that all contexts of ordinary moral practice are emotionally engaged or engaging *to some extent*, it is reasonable to suppose that the error theorist's level of emotional engagement and the force of her emotional responses in such contexts is not always sufficient to silence her belief that error theory is true. It is a familiar fact that not all recognizably moral issues and debates engage us (whether at an emotional or a philosophical level) to the same extent. Most of us have some moral values or concerns that are closer to heart than others. An ethical vegetarian, for example, might experience a higher level of emotional engagement, and stronger emotional reactions, when hearing a news report about a farmer who has been arrested on charges of systematic animal abuse than when hearing gossip about a friend's shoplifting, despite believing that both animal abuse and shop lifting is morally wrong. A libertarian might react stronger when hearing about the government's plan to increase inheritance tax than when hearing about animal-abuse, despite believing that both things are morally wrong. And so on.

Second, even given that the error theorist's emotions sometimes silences her belief that error theory is true, there seems to be a risk that she will learn, over time, not to let herself be so taken in or seduced. Someone who is repeatedly deceived by a manipulative partner might learn, over time, to brace herself and take control of her emotional responses to his suggestive behavior so that she does not let herself be seduced into believing what he says. Even if she does not actively take measures to brace herself, his ability to seduce her into believing what he says might diminish as time passes, and (if all goes well) ultimately vanish. Someone who visits a political rally for the first time in their life might find the politician's rhetorical skills and the reactions of the audience more seductive, and might easier be taken in by what the politician says, than someone who visits such meetings on a monthly or weekly basis. It seems plausible that when an agent becomes familiar with a particular context, its seductive effect wears off, even when the context is intended to put her under strong pressure to believe something.

It may be objected that there are pragmatic reasons to allow oneself to be seduced into holding genuine moral beliefs, and that this fact counteracts that successive crippling of morality's seductiveness mentioned in the previous

paragraph.²⁴ This, if correct, might speak in favor of a *hybrid theory* where conservatism and fictionalism co-exists, and where it does not matter whether the agent really accepts the content of her moral judgments or merely role-plays such acceptance, and where the agent may slide from one state to the other without great consequence. Even from the fictionalist's point of view, there need not be any harm in sometimes temporarily forgetting that one is engaged with a fiction. (More on this in chapter 8, below.)

If one is worried that that attending to the belief that error theory is true undermines moral motivation, one has to take seriously the risk that it is hard to repeatedly and reliably let oneself be seduced by the moral illusions. It seems to be a better option not to make moral motivation dependent on agents allowing themselves to be repeatedly seduced. One such option is fictionalism. According to fictionalism, the error theorist need not allow herself to repeatedly be seduced into believing things she is disposed to disbelieve in critical contexts in order to save moral motivation, for it is sufficient that she *pretends* – or, as I suggest, *role-plays* – to hold moral beliefs. When role-playing herself as moralizer, the error theorist can be aware that error theory is true, that she is engaged with a fiction, etc., and still be motivated to act according to the moral norms. In the next chapter, I spell out the details of this suggestion.

7.5 Conclusion

The topic of this chapter and the next is the practical implications of ERROR THEORY for ordinary everyday moral thinking and discourse. More specifically, the central question of these chapters is what we ought to do if we find out that ERROR THEORY is true. As I understand it, this question concerns what course of action best satisfies our shared preferences if we find out that ERROR THEORY is true. I have suggested that there are four competing answers to it, represented by four competing theories – viz. what I have referred to as *abolitionism*, *propagandism*, *conservatism* and *fictionalism*. This chapter has discussed abolitionism, propagandism and conservatism. In the next chapter, I will defend a version of fictionalism.

I began by suggesting that abolitionism seems to come at a cost, because moral practice seems to be overall (non-morally) valuable to us. It seems to provide some pretty serious benefits throughout ordinary life. Although it is a complicated empirical issue whether morality is overall beneficial to us or – as some abolitionists have suggested – overall harmful to us, I proceeded under the assumption that present moral practice is overall (non-morally) valuable.

Second, I discussed propagandism. I argued that propagandism does not provide a complete response to the relevant question, and that there seems to be

²⁴ Note that this is consistent with the truth of the claim that we typically have hypothetical reasons to care about truth and truth-conduciveness and thus have hypothetical reasons to evaluate doxastic states and actions in relation to these goals, which I defended in chapter 5.

no reason to prefer it to competing strategies the implementation of which would also leave moral practice more or less as it was prior to the discovery that ERROR THEORY is true, like conservationism.

Third, and last, I discussed conservationism. After pointing out several similarities between conservationism and fictionalism, I argued that conservationism has not given us sufficient reasons to prefer it over fictionalism, since it has not provided us with sufficient reasons to think that implementing it is better as a means of preserving benefits of moral practice than implementing fictionalism.

Chapter 8

In defense of moral fictionalism

In the previous chapter, I suggested that ordinary (first-order) moral thinking and discourse seems to be overall (non-morally) valuable to us in that it seems to provide some pretty serious benefits throughout ordinary life. The aim of the present chapter is to defend two claims. The first of these is that if we find out that ERROR THEORY is true, we can preserve these benefits by role-playing morality on the model of role-playing in *live action role-playing games* (or “LARPs”). The second claim I defend is that if moral fictionalism (from now on “fictionalism”) incorporates the suggestion that engagement in revised moral practice is modeled on role-playing in LARPs, it has potential not only to preserve benefits of present moral practice but also to handle some *prima facie* problems for fictionalism. These *prima facie* problems are that it is unclear whether engagement with the moral fiction can have bearing on motivation and action, and whether motivation can be sustained with awareness that it is guided by fiction. It is concluded that if present moral practice is overall non-morally valuable, we ought to role-play morality if we find out that error theory is true.

I proceed as follows: Section 8.1 gives an account of fictionalism, which draws upon that given by Richard Joyce (2001, 2005). Section 8.2 describes two *prima facie* problems for fictionalism. Both of these concern fictionalism’s potential of giving an account of moral motivation. Section 8.3 gives a brief description of LARPs, and explains how fictionalism can incorporate the suggestion that engagement in revised moral practice is modeled on role-playing in LARPs. Section 8.4 argues that a version of fictionalism which incorporates this suggestion has potential to handle the *prima facie* problems. I conclude that if we find out that ERROR THEORY is true, we ought to implement such a version of fictionalism; we ought to role-play morality.

8.1 Fictionalism

Fictionalism comes in many varieties.¹ The type of fictionalism that will be discussed here is a version of what is sometimes referred to as *revisionary force fictionalism*. It says that if we find out that error theory is true, we ought to carry on with moral thought and discourse as a fiction, and that doing so involves *revising* present moral practice so that it no longer involves any genuine moral beliefs or assertions (Joyce 2001:185, 2005:288, 291-6). Both Joyce's (2001, 2005) version of fictionalism and mine, which is based on his, are versions of revisionary force fictionalism (from now on, "fictionalism" refers to revisionary force fictionalism, unless otherwise indicated).² The central difference between his version of fictionalism and mine concerns our respective suggestions as to how ordinary moral practice after the implementation of fictionalism (from now on "revised moral practice") is to be carried out. I suggest that engagement in revised moral practice is to be modeled on role-playing in LARPs. Joyce does not.

In the words of Joyce, to be a fictionalist is to believe that error theory is true but to "carry on employing the discourse, at least in many contexts, as if this were not the case, because it is useful to do so (2001:185)." Just like propagandism and conservatism, fictionalism attempts to preserve benefits of present moral thinking and discourse. And just like conservatism, fictionalism recommends us to continue thinking moralized thoughts and continue using moral language (i.e., continue uttering unembedded moral sentences in the declarative.) But unlike propagandism and conservatism, fictionalism claims that we can preserve benefits of present moral practice without leaving it (more or less) as it is, i.e., without holding on to our untrue

¹ (Moral) fictionalism can be offered either as a *description* of moral practice (*hermeneutic fictionalism*), or as a proposal on how moral practice *is to be revised* (*revisionary fictionalism*). This distinction was introduced by John P. Burgess (1983). It is possible that hermeneutic fictionalism and error theory suggest mutually exclusive analyses of ordinary moral practice. That is; depending on how the hermeneutic fictionalist understands fictional practice, it is possible that if ordinary moral practice *already is* fictional practice, it does not incorporate the error(s) that the error theorist takes claims it to incorporate. For a defense of hermeneutic moral fictionalism, see Kalderon (2005b). Depending on how the revisionary fictionalist conceives of the difference between ordinary moral practice and fictional practice, there are different options as to *what kind of* revision she may recommend. The (revisionary) *force* fictionalist suggests that the force with which moral sentences are uttered is to be revised, so that the content expressed is *not asserted* (Joyce 2001:xi). In contrast, the (revisionary) *content* fictionalist suggests that the content of moral sentences is to be revised, so that they can be used to assert some *new* content. She might, for example, suggest that a (tacit) story operator is added to the content of moral sentences, so that "it is wrong to torture babies" expresses the same content as "according to the moral fiction, it is wrong to torture babies." This possible version of revisionary content fictionalism is inspired by David Lewis' (1978) account of fictional practice. For a general discussion of fictionalism and its different varieties in several different fields, see Eklund (2017).

² I find revisionary force fictionalism more plausible than revisionary content fictionalism because I believe that a plausible fictionalism should tie into our ordinary fictional competences. However, because a serious discussion of what view is in general more plausible would lead too far, and a parallel consideration of both views would complicate the discussion, my concern here will be revisionary force fictionalism.

moral beliefs, continuing to assert untrue moral propositions, or duping others into doing so. According to fictionalism, the best way to preserve benefits of present moral practice is to revise it in two related ways claims. More specifically, fictionalism recommends that we change our psychological attitudes to moral propositions and our way of using moral sentences so that we no longer genuinely believe or assert moral propositions, but only *pretend* to believe and assert them.

Fictionalism says that if fictionalism has been universally implemented (i.e. if present moral practice has universally been revised in the way fictionalism recommends), engagement in moral conversation, argumentation, thinking, reasoning, and so on *is* engagement with a fiction. If engagement in revised moral practice is engagement with a fiction, engagement in it is analogous to the activities of telling stories, performing stage plays or, as I will suggest, *role-playing in LARPs* (Joyce 2001:197-9, 2005:291).

Let us start with the recommendation that we change our psychological attitudes to moral propositions. Suppose, for example, that Stephanie is telling a story about Superman. When doing so, Stephanie pretends (or make-believes) to have access to a special realm of facts about Superman's personality, history, friends, etc. She also pretends to hold beliefs about these things. Stephanie does not believe, for example, that there once was a man who was faster than a speeding bullet. But when telling the story, she pretends to do so. She thinks, speaks and acts *as if* she held this belief.³ If Stephanie is playing the role of Superman in a Superman stage play, she moreover pretends to *be* Superman. When playing out a scene where Superman is having coffee with Lois Lane at *The Daily Planet*, for example, Stephanie might think to herself "Lois doesn't know that I am faster than a speeding bullet". But in doing so, she does not *believe* that she herself is faster than a speeding bullet, but merely *pretends to believe* this. These features of Stephanie's behavior are essential for its being the case that she is telling a story or performing a stage play, and is not describing, explaining, or talking about a story or stage play (Joyce 2005:291).⁴

Fictionalism's recommendation that we change our psychological attitudes to moral propositions is the recommendation that we take the same attitude to moral propositions as Stephanie takes to those propositions about fictional characters, history, events, etc. that make up the Superman fiction when she is telling a story about Superman or performing in a Superman stage play. That is: instead of genuinely believing moral propositions, we *pretense-believe* them.

³ Joyce (2001:197) describes make-believing as a matter of "thinking propositions" or "entertaining thoughts."

⁴ According to Joyce (2001:199-200, 2005:291), this feature of story-telling (or acting) poses a problem for content varieties of fictionalism. If every sentence uttered is prefixed by a (tacit) story-operator, there is no sense to be made of the idea that the storyteller is pretending. How could one *pretend* that according to the Siegel and Shuster story, there once was a man who was faster than a speeding bullet?

Let us now consider the recommendation that we change our way of using moral sentences. Fictionalism says that one and the same sentence can be used either to assert a proposition, or to pretend (or make-believe) to assert it, and that the difference between these two uses of a sentence is that the sentence is uttered with different kinds of *force*. Given this, we can change our way of using moral sentences without changing their contents (Joyce 2001:202, 2005:293-4, 296). Consider the following sentence

- (g) There once was a man who was faster than a speeding bullet.

If Stephanie utters this sentence in the context of talking about the real world, she asserts the proposition that *there once was a man who was faster than a speeding bullet*. Because there never was such a man, this proposition is false. So when Stephanie utters (g) in this context, she asserts something false. Suppose now that Stephanie utters (g) in the context of telling a story about Superman. Because this sentence has the same content in both contexts, her utterance expresses the same proposition in both contexts. However, when uttering (g) in the context of telling a story about Superman, Stephanie does not assert this proposition. Instead, she pretends (or make-believes) to assert it. She acts *as if* asserting it.⁵ So when Stephanie utters (g) in this context, she does not assert anything, and so does not assert anything false.

Suppose that (part of what it is) to assert a proposition, *p*, is to present *p* as something one believes and to put *p* forward as something one's audience should believe. If so, to pretend to assert *p* is to pretend to do these things – to pretend to present *p* as something one believes, and to pretend to put *p* forward as something one's audience should believe (Joyce 2005:291, 293-4).

Fictionalism's recommendation that we change our way of using moral sentences is the recommendation that we use moral sentences in the same way Stephanie uses (g) when she utters it as part of telling a story about Superman. That is: instead of genuinely asserting moral propositions, we *pretense-assert* them (Joyce 2001:199-202). More specifically: when we utter an unembedded moral sentence in the indicative, such as

- (h) it is wrong to torture animals,

we *pretense-assert* the (untrue) moral proposition that *it is wrong to torture animals*.

Whether an utterance is an assertion or a pretense-assertion seems to be determined by a framework of (linguistic and non-linguistic) *conventions* within which the utterance occurs (Joyce 2001:203). Consider sarcasm. There is a

⁵ In the words of John R. Searle (1975:320-1), Stephanie is *not seriously committed* to the proposition expressed.

widespread convention which says that if one utters the sentence “That dinner party was fun” in a certain tone of voice (a tone of voice dripping with sarcasm), one does not thereby assert that the event was fun (Joyce 2001:11). Similarly, there is a widespread convention stipulating that if an utterance is preceded by “once upon a time”, the speaker may not believe that which follows, and is not putting it forward for others to believe (Joyce 2001:12).

According to Joyce, it is possible for us to start using moral sentences in the way we use sentences that we utter as part of telling a stories or participating in stage plays, because

[...] a group of speakers could choose to exploit extant conventions, or develop new ones, such that an understanding develops that when one enters into moral discourse one is performing a linguistic activity the sentences of which will not (generally) be endorsed as true in very critical environments, and therefore do not count as belief-expressing (Joyce 2001:204).

It is important to remember that fictionalism is suggested as a strategy that a *group* of people might implement if they find out that error theory is true. For while it seems possible for a group to start using moral discourse in a non-assertoric manner, the suggestion that a particular utterance’s status as assertion or pretense-assertion is determined by within what framework of conventions it is made seems to imply that it is at least improbable that a single individual could pull this off by themselves (see Joyce 2001:203-4, 2005:296-8 and Garner 2010:227-9 for discussion.)

Some have suggested that whether an utterance is a pretense-assertion depends on the speaker’s intentions. John R. Searle, for example, seems to suggest this in the following passage:

One cannot truly be said to have pretended to do something unless one intended to pretend to do it. So [...] the identifying criterion for whether or not a text is a work of fiction must of necessity lie in the illocutionary intentions of the author. [...] What makes it a work of fiction is [...] the illocutionary stance that the author takes toward it, and that stance is a matter of the complex illocutionary intentions that the author has when he writes or otherwise composes it (Searle 1975:325).

It seems correct that a speaker can utter a sentence with the *intention* that it has one force rather than another. But one might think that context can determine force, so that someone who *misunderstands* the context can make an assertion unintentionally. Suppose, for illustration, that Miranda has implemented

fictionalism because she has found out that error theory is true – i.e., that she is a *fictionalizer*.⁶ Suppose that Miranda utters (h) with the intention to use fictional force among people who – unbeknownst to her – are *not* fictionalizers, and who have a convention which says that when one utters an unembedded moral sentence in the indicative, using a serious tone of voice, showing a face expression and using the kind of gestures that seem normal and expected given the circumstances, without preceding the sentence with something equivalent to “once upon a time...”, etc., one is to be taken as having made an assertion. Suppose also that Miranda does not do anything to signal that she does not intend to use assertoric force in uttering (h). If so, it is possible that her audience interprets her as *asserting* that it is wrong to torture animals. It is also possible that if her audience interprets her utterance as an assertion, *it is an assertion* regardless of whether Miranda intended to use assertoric force. Given this, it seems that the speaker’s intentions do not wholly determine whether her utterances are pretense-assertions or genuine assertions (see Joyce 2001:12, 2005:296-7).

The question of to what extent, if any, speaker’s intentions determine whether an utterance is a pretense-assertion or not is a difficult one. The same goes for the question of how linguistic conventions are established and passed on. Fortunately, we need not decide on these matters for the purposes of the discussion of this chapter. The important point is that it is possible for a group of speakers to have a convention in place that when (in contexts of ordinary conversation) the subject matter of morality is entered into, and someone makes a moral utterance with the intention to pretense-assert the proposition constituting the content of the sentence used, she does not assert this proposition but merely pretense-asserts it (see Joyce 2005:297-8).

In this section, I have spelled out fictionalism’s central claim that if we find out that error theory is true, we ought to revise present moral practice so that it no longer involves any genuine moral beliefs or assertions, but only pretense-assertions and pretense-beliefs. To pretense-believe or pretense-assert a moral proposition is to act *as if* one believed that there are moral facts and properties, non-institutional desire-transcendent reasons, irreducibly normative favoring relations, etc. The claim that we ought to revise moral practice in this way if we find out that error theory is true is motivated by the idea that doing so is a way of preserving (at least some) benefits of present moral practice while avoiding untrue moral beliefs and assertions. Given this, the ultimate aim of pretense-assertion must be to induce pretense-beliefs, pretense-judgments, and genuine desires and emotions that might influence actions and behavior in one’s audience, without thereby inducing any genuine moral beliefs in them, and without leading them to make any genuine moral judgments or assertions.

⁶ I use “fictionalizer” to refer to someone who – like Miranda – has implemented fictionalism after finding out that error theory is true. In contrast, I use “fictionalist” to refer to someone who defends or accepts some sort of fictionalism. Fictionalists may or may not be fictionalizers.

8.2 Two *prima facie* problems for fictionalism

In this section, I describe two *prima facie* problems for fictionalism; in the two sections that follow, I argue that a version of fictionalism which models engagement in revised moral practice on role-playing in *live action role-playing games* has potential to handle these problems.

The first *prima facie* problem for fictionalism is that it is unclear whether engagement with the moral fiction can have bearing on motivation and action. One of the considerations offered in support of fictionalism is that it is a means of preserving benefits of present moral practice. I suggested above that it seems plausible that moralized thoughts have some bearing on motivation and action regardless of whether the agent holds any genuine moral beliefs. But it can be argued that part of the reason why present moral practice is beneficial for social cooperation, counteracting limited human sympathies, etc., is that there is an intimate connection between moral belief, on the one hand, and motivation and action, on the other.⁷ It is often suggested that our moral beliefs function as a check on improper motivation – i.e., that they keep us from acting on some of our desires (Hussain 2010:341). If this is the case, and it turns out that moral pretense-beliefs cannot play this functional role, implementing fictionalism does not seem like a fruitful strategy for preserving benefits of present moral practice.

In defense of the claim that there is a causal link between engagement with the moral fiction and action, Joyce argues that:

Reading *Anna Karenina* may encourage a person to abandon a doomed love affair; watching *The Blair Witch Project* may lead one to cancel the planned camping trip in the woods. [...] [T]hese aren't the kind of beneficial behavioral responses that the moral fictionalist is seeking, but they at least show that the causal links between involvement with a fiction and action are undeniably in place (Joyce 2005:303).

Joyce suggests that in analogy to how engagement with Tolstoy's literary fiction may encourage someone to abandon a doomed love affair, engagement with the moral fiction may encourage someone to, for example, remain truthful to their spouses. In response to this, Olson argues that it is unclear whether engagement with the moral fiction can have bearing on motivation and action without causing *genuine* moral beliefs:

Upon reading *Anna Karenina* I may come to believe that a love affair recently embarked on is hopeless; upon watching *The Blair Witch Project* I may come to believe that [...] a crazy serial

⁷ This connection may or may not be internal or necessary, as motivational internalists claim.

killer roams the woods. And these beliefs about reality [...] may serve as partial explanations of subsequent behaviour (Olson 2011*b*:195, 2014:187).

The suggestion that engagement with the moral fiction can motivate us to act only by giving rise to genuine moral beliefs creates two problems for fictionalism. First, moral pretense-beliefs cannot in themselves motivate us into action. Second, contrary to what fictionalists suggest, implementing fictionalism is not a way of avoiding genuine moral beliefs (Olson 2011*b*:195, 2014:187).

Another reason one might have for doubting that engagement with the moral fiction would have bearing on motivation and action has been pointed out by Graham Oddie and Daniel Demetriou. They suggest that “make-belief is [...] a highly overridable attitude (Oddie and Demetriou 2010:200)”, and that “whenever we make-believe something we are primed to abandon the attitude if reality intrudes in a rude or demanding way (ibid).” They defend this claim by drawing an analogy between someone watching a stage play and a fictionalizer (i.e., someone who, after finding out that error theory is true, has implemented fictionalism). Suppose that the theatergoer is sitting comfortably in his seat when a fire breaks out backstage. Oddie and Demetriou claim that

[...] if smoke starts seeping onto the stage from backstage but it is clearly true, in the play, that there is no smoke in the room, we tend [...] to abandon the make-belief (that there is no smoke in the room) and go with the belief (that there is smoke), and it is entirely reasonable to do so. When push comes to practical shove, make-belief will rightly give way to genuine belief (ibid:201).

Consider now the fictionalizer who pretends to believe that lying is wrong. Because she pretends to hold this belief, she (among other things) generally avoids lying. But why, ask Oddie and Demetriou, wouldn't she have more reason to be guided by her genuine beliefs about lying (i.e., her second-order belief that it is not the case that lying is wrong) in circumstances where doing so serves her rude self-interest? The intended answer seems to be that the fictionalizer will abandon her pretense-beliefs in such situations (ibid).

The second *prima facie* problem for fictionalism is that even if engagement with the moral fiction can have bearing on motivation and action, it is unclear whether motivation can be sustained with awareness that it is guided by fiction. We saw above that fictionalism recommends a moral compartmentalization. More specifically, it recommends the following: In the uncritical contexts of revised moral practice, the fictionalizer immerses herself in the moral fiction and pretends to believe and assert moral propositions. In critical contexts, she attends to her genuine beliefs about morality – that error

theory is true, that nothing is morally right or wrong, that her engagement in ordinary moral practice is engagement with a fiction, etc. – and does not immerse herself in the moral fiction. We have also seen that one of the considerations offered in support of fictionalism is that it is a means of avoiding genuine moral beliefs. This gives rise to a potential problem for fictionalism, for it can be argued that in order to avoid genuine moral beliefs, the fictionalizer will sometimes have to attend to her genuine beliefs about morality even when engaged in revised moral practice. And it can moreover be argued that if the fictionalizer can be aware of her genuine moral beliefs when engaging in revised moral practice, the first problem reappears: it is unclear why she would be *significantly motivated* – i.e. motivated in such a way that her pretense-beliefs function as a check on improper motivation – while aware that her engagement with the moral fiction, and her moral pretense-beliefs, is mere pretense.

According to Olson, this reveals a practical tension in fictionalism. Olson raises this worry in response to Joyce's claim that

[t]he decision to adopt morality as a fiction is best thought of as a kind of precommitment. [...] [T]he resolution [...] is something that occurred in the person's past, and is now an accustomed way of thinking. Its role is that when entering a shop, the possibility of stealing doesn't even enter her mind. [...] What goes through her mind may be exactly the same as what goes through the mind of the sincere believer – it need not 'feel' like make-believe at all [...]. The difference between the two need only be a disposition that the fictionalist has (though is not paying attention to): the disposition to deny that anything is really morally wrong, when placed in her most critical context (Joyce 2005:306).⁸

As Olson (2011b:197, 2014:189) points out, Joyce seems to suggest that to implement fictionalism is to adopt certain behavior-patterns and ways of thinking (i.e., thinking moralized thoughts) in uncritical contexts while being disposed to deny that anything is really right, wrong or forbidden in critical contexts.

Olson defends the claim that there is a practical tension in fictionalism by arguing as follows: First, the fictionalizer engaged in revised moral practice will have to be on her guard not to slip from holding pretense-beliefs into holding genuine moral beliefs (because it will be difficult for someone who has successfully adopted the relevant behavior-patterns and ways of thinking to avoid forming genuine moral beliefs). And practicing such self-surveillance seems to involve occasionally attending to one's genuine belief that morality is a

⁸ Joyce uses "fictionalist" in this passage to refer to someone I would call a "fictionalizer", i.e., to someone who, upon finding out that error theory is true, has implemented fictionalism.

fiction. Second, attending to this belief reduces the motivational effectiveness of one's engagement with the fiction:

[...] in order for moral precommitments to be effective in bolstering self-control, beliefs to the effect that morality is fiction need to be suppressed or silenced (Olson 2011b:197, 2014:189).

According to Olson, fictionalism's practical recommendations pull in opposite directions, because it is unclear whether it recommends the fictionalizer to (i) occasionally attend to, or to (ii) suppress her genuine belief that morality is a fiction (Olson 2011b:197-8, 2014:189-90).

Köhler and Ridge suggest that because attending to one's genuine beliefs about morality reduces the motivational effectiveness of one's engagement with the fiction, it is easy for the fictionalizer to take steps to undermine the motivational effectiveness of her moral pretense-beliefs when there are personal gains in sight. She may simply slide into the critical perspective:

After all, when adopting the critical perspective we are aware that our [...] [everyday moral] thoughts are just engagements with a fiction, which should plausibly diminish how serious we take them (Köhler and Ridge 2013:15).

If the fictionalizer will take steps to undermine the motivational effectiveness of her moral pretense-beliefs – or altogether abandon them, as Oddie and Demetriou (2010) suggest – when there are personal gains in sight, moral pretense-beliefs cannot serve the function of keeping improper motivation in check.

In this section, I have described two *prima facie* problems for fictionalism. The first is that it is unclear whether engagement with the moral fiction can have bearing on motivation and action. Some argue that engagement with the moral fiction can motivate us to act only by giving rise to *genuine* moral beliefs, others that we are primed to abandon the attitude of make-belief if reality intrudes in a rude or demanding way. The second *prima facie* problem for fictionalism is that even if engagement with the moral fiction can have bearing on motivation and action, it is unclear whether motivation can be sustained with awareness that it is guided by fiction. Some argue that there is a practical tension in fictionalism, because while the fictionalizer engaged in revised moral practice will sometimes have to attend to her genuine beliefs about morality, attending to these beliefs undermines the motivational effectiveness of her moral pretense-beliefs. Others argue that it will be easy for the fictionalizer to take steps to undermine the motivational effectiveness of her moral pretense-beliefs when there are personal gains in sight – she may simply

slide into the critical perspective. In the following two sections, I spell out a version of fictionalism which has the potential to handle these problems.

8.3 Modeling revised moral practice on LARPs

This section gives an account of how a fictionalism of the kind described above (section 8.1) can incorporate the suggestion that engagement in revised moral practice is modeled on role-playing in *live action role-playing games* (from now on, “LARPs”) continuously running in parallel with the player’s ordinary lives.

LARPs can be introduced as fusions of improvisational theatre and rule governed games. Except for an overall frame story, LARPs usually have no predetermined content or plot. A player participating in a LARP uses her imagination, creativity and improvisation to develop and physically portray – that is, *role-play* – a fictional character. She improvises the feelings, utterances, behavior, and actions of her character (however, characters usually have a set of predetermined character traits). During the game, the player’s actions and utterances represent actions and utterances of her character. In this way, players jointly create the story of the LARP. Two other central aspects of LARPs is that LARP-players often spend weeks preparing and manufacturing elaborate costumes and props to use in a LARP, and that objects figuring in the physical environment in which a LARP is played, might represent other objects in the game – a silicone sword might represent a steel sword covered with diamonds, a string of twine on the floor might represent a magical force field, and so on (Harding 2007:25-8).

In addition to the central roles of imagination, creativity and improvisation, LARPs and improvisational theater can have similar purposes. The purpose of LARPs is often described in terms of interaction, cooperation (e.g. the collaborative creation of a story), challenge (e.g. overcoming challenges or obstacles in the course of pursuing the character’s projects and aims), enjoyment, personal development, and deepened emotional self-knowledge. One difference between LARPs and improvisational theatre is that LARPs are not directed towards, or attempting to achieve some effect within, anyone but the players. Another is that LARPs are geographically and temporally unlimited to a greater extent than stage plays. While a stage play is usually performed on a theater stage and lasts for no longer than an evening, a LARP may take place in a variety of physical settings (but usually not on a theater stage) and last for several days or even years, with the game played in parallel with the players’ ordinary lives. An important difference between LARPs and rule governed games is that LARPs usually have no winner or loser.

“Immersion” – as I use it – picks out the kind of pretense or make-belief involved in LARPs, and/or the state of mind of an agent engaged in such

pretense.⁹ It can be characterized roughly as follows: Suppose that Stephen – an ordinary human being – is participating in a LARP, and is role-playing a 505 year old cyborg whose strongest desire is to revenge its father, and who believes that its father was murdered by a group of red-bearded men. When Stephen immerses in the cyborg, he is in a state of mind where he does not need to actively suspend disbelief, and role-plays the cyborg in a way that feels “natural” and easy to him. Stephen uses his imaginative, creative and deliberative capacities to think up what feelings, desires, beliefs, etc. the cyborg would have in situations played out in the LARP. He pretends to have these feelings and desires, and to hold these beliefs. He then acts as if he held these feelings, desires and beliefs. In doing this, Stephen focuses his attention on the cyborg’s feelings, desires and beliefs – i.e. on his pretense-feelings, pretense-desires, and pretense-beliefs – and less on his genuine feelings, desires, and beliefs. (For discussion, see Kim 2004:37; Utne 2005:25-6; Harding 2007:30-1; Holter 2007:19-21.)

The kind of fictionalism defended here is similar to that defended by Joyce in that it says that if we find out that error theory is true, we ought to continue with morality as a fiction, and that doing so centrally involves revising present moral practice so that engagement in it no longer involves holding genuine moral beliefs and making genuine assertions but merely involves holding moral pretense-beliefs and making pretense-assertions. It differs from Joyce’s fictionalism in that it suggests that the kind of pretence involved in revised moral practice is *role-playing*. According to my suggestion, engagement in revised moral practice is analogous to role-playing in a LARP continuously running in parallel with the players’ ordinary lives. This suggestion should be understood as entailing that agents engaged in revised moral practice *role-play morality by role-playing themselves as moralizers*. More specifically, agents role-play themselves holding those very moral beliefs, and making those very moral judgments and assertions which (they imagine that) they themselves would have held and made if they had never found out that error theory is true.

Suppose, for illustration, that Chad is part of a community of speakers who, after realizing that error theory is true, has started role-playing morality. In ordinary moral contexts, Chad role-plays himself as believing that there are moral facts and properties, non-institutional desire-transcendent reasons, irreducibly normative favoring relations (or whatever error theory claims to be presupposed in present moral practice, see below), as making judgments and assertions, and as reasoning and arguing in ways that imply or presuppose that this is the case. The character he pretends to be, and immerses in, is himself holding moral beliefs, making moral judgments and assertions, and reasoning and arguing morally. Chad uses his imaginative, creative and deliberative

⁹ I have no intention to take sides in the debate, within the LARP-community, over how “immersion” is best understood.

capacities to think up what moral beliefs he would have had, what moral judgments and assertions he would have made, and how he would have reasoned morally in specific (actual or hypothetical) contexts *if* he had believed that there are moral facts and properties, (etc.). When doing this, Chad gives moral issues serious and careful consideration, and lets the conclusions of this (moral) deliberation guide his actions in morally relevant situations. Just like how Stephen pretense-believes and pretense-asserts propositions about cyborgs, groups of red-bearded men, and other things that make up the fiction of the LARP with which he is engaged, Chad makes moral pretense-judgments, and pretense-believes and pretense-asserts propositions about moral facts, moral properties, and other things that make up the *moral* fiction. And just like Stephen, Chad focuses his attention on his moral pretense-beliefs and moral pretense-judgments and less on his genuine beliefs about morality (e.g. his second-order belief that nothing is morally wrong because error theory is true).

Because Chad is role-playing himself as (he imagines that) he would have been if he had never found out that error theory is true, the contents of the specific moral beliefs, presuppositions, judgments and assertions he pretends to hold and make when role-playing himself as moralizer is mainly determined by the contents of the moral beliefs, judgments and assertions he held or made when engaging in first-order moral thinking and discourse prior to realizing that error theory is true. If he believed that it is wrong to torture animals, he pretends to believe that it is wrong to torture animals. If he was prone to judge, in most situations, that others morally ought to be truthful to their spouses, he pretends to be prone to judge, in most situations, that others morally ought to be truthful to their spouses. And so on.

This being said, the contents of the beliefs, presuppositions, and so on, Chad pretends to hold and make when role-playing himself as moralizer need not be wholly determined by those (positive) first-order moral beliefs, presuppositions, and so on, he held and made prior to realizing that error theory is true. They may also partially be determined by the nature of the argument that convinced him, or made him realize, that error theory is true. Some of the beliefs he pretends to hold when role-playing morality may, in other words, be of the second-order (metaethical) kind – i.e. be beliefs *about* first-order moral beliefs, judgments, reasoning, and language.

As suggested in earlier chapters, different arguments can be offered for error theory. These arguments represent different grounds based on which someone may come to believe or realize that error theory is true. The argument from non-institutionality, for example, says that morality involves an error because (present) moral discourse presupposes non-institutional desire-transcendent reasons and non-institutional categorical imperatives (see chapter 3). If Chad comes to believe that error theory is true on the basis of this argument, he probably believes its premises. If he does, he believes that moral discourse presupposes non-institutional desire-transcendent reasons and non-

institutional categorical imperatives. When role-playing himself as moralizer, he will for this reason pretend to believe or presuppose that there can be non-institutional categorical requirements on agents to do something, so that agents can have non-institutional desire-transcendent reasons for action. The moral fiction Chad immerses in when role-playing morality can be silent with respect to all metaethical issues except for those very metaethical claims that figure in that argument on the grounds of which he accepts error theory.

The fictionalism defended here is similar to Joyce's fictionalism in that it recommends that we role-play ourselves as moralizers in everyday contexts of revised (first-order) moral practice – i.e., in precisely those contexts where we, prior to realizing that error theory is true, would make genuine positive first-order moral judgments, genuinely assert positive first-order moral propositions, and attend to our genuine positive first-order moral beliefs. It does not recommend that we role-play ourselves as moralizers in contexts where we are engaged in critical thinking and serious contemplation of metaethical issues, such as the philosophy seminar room. It does not recommend, that is, that we role-play ourselves as moralizers when discussing morality from a *metaethical* perspective. The agent who role-plays herself as moralizer when discussing the ethics of abortion with a devout catholic, when trying to determine whether to vote in the church election, or when teaching her kids about what it is to be a good friend, need not role-play when attending a philosophy seminar on the first chapter of Mackie's *Ethics*, when discussing the issue of whether moral language is essentially descriptive, or trying to determine how the metaphysics of moral facts are best described. The latter are, and the former are not, contexts where she is thinking critically and seriously contemplates metaethical issues. However, the fictionalism defended here differs from Joyce's fictionalism in that it does not recommend a moral compartmentalization or two-level thinking, but allows agents engaged in revised moral practice to attend to their beliefs that error theory is true, that they are engaged with a fiction, and so on.¹⁰

In this section, I have spelled out my suggestion that if we find out that error theory is true, we ought to revise present moral practice on the model of role-playing in a LARP continuously running in parallel with the player's ordinary lives. An agent engaging in revised moral practice modeled on such role-playing role-plays herself as moralizer – i.e., role-plays herself holding those very moral beliefs, and making those very moral judgments and assertions which (she imagines that) she herself would have held and made if she had never found out that error theory is true. In doing so, she gives moral issues serious and careful consideration, and lets the conclusions of such (moral) deliberation guide her actions in morally relevant situations. The *contexts* in which she role-plays herself as moralizer are precisely those where she, prior to realizing that error theory is true, made genuine first-order moral judgments,

¹⁰ For more on moral compartmentalization, see chapter 7.

genuinely asserted first-order moral propositions, attended to her genuine first-order moral beliefs, and so on. The contents of the specific moral judgments, assertions, and beliefs this fictionalizer pretends to make and hold is mainly, but not wholly, determined by what first-order moral judgments, assertions and beliefs she held and made prior to realizing that error theory is true. In the next section, I argue that when fictionalism is understood as incorporating the suggestion that engagement in revised moral practice is modeled on role-playing in LARPs, it has potential to handle the *prima facie* problems for fictionalism spelled out in section 8.2, above.

8.4 Handling the *prima facie* problems

This section argues that a version of fictionalism which models engagement in revised moral practice on role-playing in LARPs can handle the *prima facie* problems for fictionalism described above (section 8.2) – viz. that it is unclear whether engagement with the moral fiction can have bearing on motivation and action and whether motivation can be sustained with awareness that it is guided by fiction.

Both problems concern fictionalism's potential of giving an account of *moral motivation*. But what kind of moral motivation are we talking about here? The answer to this question seems to be determined by what kind of moral motivation is delivered by present moral practice. I claimed above that there seems to be an intimate connection between moral belief and motivation. Common experience – that is, our experience with present moral practice – seems to suggest that our moral judgments typically motivate us, at least to some degree.¹¹ But it also seems to be part of our common experience that present moral practice does not deliver an intrinsic or necessary connection between moral judgments and significant motivation. Many have pointed out that although most of us have views on what morality requires from us, and are anxious to comply with these requirements, we sometimes fail to act according to them. Sometimes, we are simply too tired to do all the things that morality requires of us. We are lazy about recycling our newspapers, and allow ourselves to capitulate on minor matters of principle, etc. Sometimes, these moral lapses do not even make us feel guilty.

Kate Manne suggests that sometimes when an agent makes a first-personal moral judgment (about what she currently ought to do), and fails to be motivated to act in accordance with this judgment, she fails to *relate to herself*

¹¹ That is, when we judge that some action would be morally right, we are typically motivated – at least to some degree – to perform this action. The nature of this connection is of course contested. There is no agreement on whether the connection between moral judgment and motivation is necessary or contingent. There is also disagreement on whether moral judgments can motivate on their own, or merely in conjunction with some (pre-existing) desire or (other) desire-like or intrinsically motivating state. I will not attempt to answer any of these questions. For more on these issues, see Rosati (2016).

as an active participant in moral practice (Manne 2015:262-3). She fails to have a sense of herself as an insider to the moral community at large (ibid:267). When this happens, says Manne,

[...] [the agent] is telling herself what to do, but responding only passively, almost as if she is overhearing a direction being issued to someone else. [...] At one level, the agent acknowledges that the moral judgment she makes properly applies to her, and it is at least open to us to take her judgment at face value. So she might say or think to herself “I really ought to ϕ ,” and this statement may be perfectly sincere and not even half-hearted, exactly. Rather, or better, it is bloodless—or, alternatively, mindless. [...] The agent’s sense of agency has somehow gone missing. She may have views about what to do, but it’s as if she forgets herself (Manne 2015:262-3).

According to Manne, the reason why we sometimes fail to relate to ourselves in this way may be that we realize that we are currently not in our right mind, or that we are not quite ourselves. It may also be that we are overwhelmed with other responsibilities, or simply lazy (ibid:269-270).

Since I have no intention to take sides in the motivational internalism versus externalism debate, nor in the debate on what constitutes the best formulation of motivational internalism, I will not go further into the details of Manne’s proposal.¹² Still, it seems plausible to me that *if* real moral judgments *can* motivate us to act, then in order for them to do so, we need to somehow view or relate to ourselves as active participants in moral practice. That is; we need to see ourselves *as occupying a specific role*. It also seems plausible to me that we can drift out of this perspective, and sometimes also do so, because we are tired, overwhelmed, lazy, etc.

If something like this is true with respect to present moral practice for roughly these reasons – i.e., if motivation by genuine moral judgments is tied to roles we can see ourselves as playing – it has three important implications for fictionalism. First: If genuine moral judgments do not *necessarily motivate* us to act according to them, it is neither fair nor reasonable to require that moral pretence-judgments necessarily do so. Second: If genuine moral judgments motivate us to act only when we see ourselves as occupying a specific role, modeling revised moral practice on LARPs captures an important feature of present moral practice, and makes the difference in how real moral judgments and moral pretence-judgments motivate less pronounced. The suggestion that

¹² Manne defends Tempered internalism – i.e. the claim that “a moral judgment made by an agent about what she herself (currently) ought to do will entail motivations on her part to act in accordance with this judgment, *provided that she takes the practical stance towards herself which is fitting for such judgments* (Manne 2015:263).”

revised moral practice is modeled on LARPs gives a clear idea of how agents engaging in revised moral practice can relate to themselves as active participants in moral practice – viz. by playing the role of themselves as moralizers. When we play the role of ourselves as moralizers, we get a sense of ourselves as insiders to the moral community, and see ourselves as active (rather than passive) and as implicated (rather than exempt) (see *ibid*:267-8). Third: Role-players engaged in LARPs may sometimes drift out of the roles. But if Manne is right – as I believe that she is – this is true also of agents engaging in present moral practice. They are only motivated by their (genuine) moral judgments if they relate to themselves as active participants in moral practice, and sometimes drift out of this perspective. So there seems to be a straightforward explanation of why agents engaged in revised moral practice modeled on role-playing in LARPs may sometimes fail to be motivated by their moral pretence-judgments.

In the remainder of this section, I will explain how the suggestion that engagement in revised moral practice is modeled on role-playing in LARPs helps fictionalism handle the *prima facie* problems discussed above.

The first problem was that *it is unclear whether engagement with the moral fiction can have bearing on motivation and action*. One reason someone might have for thinking that this is unclear, discussed above, was that it is unclear whether engagement with the moral fiction can have bearing on motivation and action *without causing genuine moral beliefs*. My response to this is that because a role-player's engagement with the fiction of a LARP can have bearing on his motivation and action without giving rise to genuine beliefs in any of the propositions that make up the relevant fiction, it is reasonable to suppose that a fictionalizer's engagement with the moral fiction can have bearing on his motivation and action without giving rise to genuine beliefs in any of the moral propositions that make up this fiction.

When someone role-plays a character that is not identical to himself in every respect (e.g. when Stephen, who is an ordinary human being, role-plays a cyborg), it is clear from the point of view of an outside spectator that there is a distinction to be made between, on the one hand, that which is true to the character but not to the role-player – i.e., true in the fiction of which the character is a part, but not true in the (real) world outside the fiction – and, on the other hand, that which is true to the role-player but not to the character – i.e., true outside the fiction, but not true in the fiction.¹³ The role-player himself is also normally able to keep track of this distinction *even when engaged in role-playing*. Moreover, his engagement with the fiction of the LARP need not – and normally does not – give rise to any genuine beliefs in any of the propositions that are true in the fiction of the LARP but not outside it.

¹³ It seems to me that this distinction is well-established within the LARP community (See e.g. <https://nordiclarp.org/wiki/Diegesis>).

Suppose that Stephen is role-playing a character that has the ability to walk through walls. Being an ordinary human being with no supernatural powers, Stephen does not have this ability. When role-playing and immersing in this character, Stephen does not genuinely believe that he can walk through walls, but merely pretends to hold this belief. Just like in his ordinary life outside the game, Stephen genuinely believes that he cannot walk through walls. If, for example, at some point during the game, his character enters a building by walking through its wall, Stephen will not attempt to role-play this event by actually walk through a real wall. He will role-play this event by some other means. This observation suggests that in this moment, Stephen attends to his belief that he cannot walk through walls. But the fact that when role-playing a character that can walk through walls, Stephen genuinely believes that he cannot walk through walls *does not imply* that he attends to this belief at all times when role-playing. If, for example, Stephen is role-playing his character enjoying an evening meal by a camp fire in the woods (far away from any walls or other things that might remind him of his belief that he cannot walk through them), he might not attend to this belief at all, but be fully focused on the feelings, desires, beliefs, etc. he imagines that his character would have in the scene presently played out. The same thing seems to be true for Stephen's ordinary life outside the game. The fact the he genuinely believes that he cannot walk through walls does not imply that he attends to this belief at all times, and so does not imply that he attends to it when, for example, taking his dog for a morning walk.

But although role-players can engage in LARPs without believing the propositions that make up the fiction of the LARP, their engagement with the fiction can have bearing on motivation and action. This point can be brought out by reflecting on what goes on when a role-player engaged in a LARP is handling an object whose real characteristics differs from its characteristics in the LARP, and her handling of this object is part of the game. For example: In LARPs that involve sword fights between characters, players use "boffer swords" – padded silicone swords – which represent real swords in the fiction of the game. Suppose that Stephen is role-playing a sword fight using a boffer sword. If so, it is true in the fiction that his character is swinging a real sword made out of steel at its antagonist, and true in the real world that Stephen is swinging a boffer sword made out of silicone at another player. Although Stephen does not believe that he is swinging a real sword made out of steel at another player, he can act as if this was what he was doing. If Stephen is physically strong and the boffer sword is light, but it is part of the fiction that his character is weak and the sword heavy, Stephen can act as if the boffer sword was heavy in his hands and as if it was difficult for him to swing it. As illustrated by the previous example, Stephen might not be attending to his belief that he is handling a boffer sword at this very moment. But it is also possible that he is aware, at least at some level, that he is handling a boffer sword. If the boffer sword has been repaired with

sticky duct tape, for example, the sticky sensation in his palms might serve to remind him that the item in his hands is in fact made out of silicone and not out of steel (duct tape rarely serves to hold a heavy steel object together). This, however, need not stop him from being immersed his character and the events presently played out in the LARP, and so need not stop his engagement with the fiction from having a bearing on his motivation and action (more on this below).

I take the above examples to illustrate two things. First, Stephen's engagement with the fiction of the LARP *can* have bearing on his motivation and action. Second, when Stephen role-plays his character, we can explain his behavior *without* assuming that his engagement with the fiction of the LARP has given rise to any genuine (false) beliefs in some of the propositions that make up this fiction. This suggests that if engagement in revised moral practice is modeled on role-playing in LARPs, the fictionalizer's engagement with the moral fiction can have bearing on her motivation and action, and when the fictionalizer role-plays herself as moralizer, her behavior can be explained without reference to any genuine (untrue) moral beliefs that her engagement with the moral fiction gives rise to.

A second reason someone might have for thinking that engagement with the moral fiction cannot have bearing on motivation and action, discussed above, was that *the fictionalizer will abandon her pretense-beliefs whenever doing so serves her self-interest*. My response to this is that the fictionalizer has non-moral reasons to role-play morality and to hold on to her moral pretense-beliefs, and that these reasons are intact in contexts when her selfish desires and interests tempt her to deflect.

One set of such reasons reveal themselves when we reflect on the considerations that motivated a group of fictionalizers to implement fictionalism and to (thereby) start role-playing morality. They chose to do so for two reasons. First, they took present moral practice to be non-morally valuable (instrumentally valuable for counteracting limited sympathies, coordinating attitudes, reflective and coherent planning in the face of temptation, individual decision making, etc.), and therefore took themselves to have something to gain from preserving those moral habits (thinking moralized thoughts, using moral language, etc.) which were in place prior to the discovery that error theory is true. Second, they had come to the conclusion that implementing fictionalism was the most efficient way of salvaging as much as possible of the valuable aspects of present moral practice. Fictionalizers thus have non-moral reasons to stick to the moral fiction. There is a set of benefits they desire to get hold of, and they believe that role-playing morality is the best way of getting hold of them.

Moreover, the individual fictionalizer's personal and selfish interests, cares and desires might provide her with additional non-moral reasons to stick to the moral fiction in the face of temptation. She might take herself as standing to gain some *personal* benefit from being (or becoming) a person who acts and reasons as if there really were moral facts and properties, non-institutional

desire-transcendent reasons, etc., even when tempted. If the fictionalizer believes that cultivating and cherishing her moral habits will help her act in ways that are, in the long run, in her own personal best interest, she has non-moral reasons to do so.

Acting according to morality can sometimes be a way of making sure that we do not hurt those we care about. If Brad cares about his girlfriend enough not to want to hurt her feelings, and believes that her feelings would be severely hurt if she caught him cheating, he has non-moral reasons to make sure that he is never caught cheating. But suppose that Brad is aware that he is epistemically fallible in the sense that even if he believes, in some situation, that he will get away with cheating, it is possible that he will actually get caught (for discussion, see Joyce 2005:300). Given this, the fact that Brad cares about his girlfriend seems to provide him with non-moral reasons to be (or become) a person who, in practical deliberation, never even considers cheating an available course of action. One way of being such a person is to reason *as if* cheating was morally forbidden, and in this way put the decision to cheat off the table in practical deliberation (see Köhler and Ridge 2013:6-7). And one way of reasoning as if cheating was morally forbidden is to role-play morality. So the fact that Brad cares about his girlfriend seems to provide him with non-moral reasons to role-play morality.

Acting according to morality can also be a way of making sure that we do not damage our own well-being. If Brad cares about his own well-being, and believes that getting caught shoplifting would have a negative effect on his future wellbeing (at the very least, because it would have a negative effect on his future economic situation), he has non-moral reasons not to get caught shoplifting. Supposing, once again, that Brad is aware of his epistemic fallibility, the fact that Brad cares about his own future well-being seems to provide him with non-moral reasons to be a person who never even considers shop-lifting an available course of action. Hence, he seems to have non-moral reasons to role-play morality.

We saw above that Oddie and Demetriou claim that the situation of the tempted fictionalizer is analogous to that of someone sitting in the audience at a burning theater – both fictionalizers and theatergoers will abandon the attitude of make-belief when push comes to practical shove. But in order for this analogy to support the claim that the tempted fictionalizer will abandon her moral pretense-beliefs, it has to be the case that *acting according to morality when one's self-interest would be better served by defiance* is for the fictionalizer what *remaining in one's seat when the theater is on fire* is for the theatergoer. However, while it is hard to come up with any reason for the theatergoer to stay in her seat – whether in virtue of his being a theatergoer, or in virtue of her being a rational agent – we have already seen that there seem to be several non-moral reasons for the fictionalizer to stick to the moral fiction, and to continue pretending to hold moral beliefs. These reasons are intact in contexts where it

would serve her self-interest to “to abandon the make-belief [...] and go with the belief (Oddie and Demetriou 2010:201).” For this reason, the analogy fails.¹⁴

I take the above reflections to suggest that fictionalizers have non-moral reasons to stick to the moral fiction and their moral pretense-beliefs even when tempted to deflect. Given this, it seems to me that, contrary to what some have suggested, fictionalizers will not abandon their pretense-beliefs whenever doing so serves their self-interest.

The second *prima facie* problem for fictionalism was that even if engagement with the moral fiction can have bearing on motivation and action, *it is unclear whether motivation can be sustained with awareness that it is guided by fiction*. We saw that it can be argued that the fictionalizer needs to suppress or silence her second-order beliefs about morality (that error theory is true, that morality is fiction, etc.) in order to make sure that the causal connection between her engagement with the fiction and action does not break down, and that Köhler and Ridge takes this to suggest that it will be easy for her to undermine the motivational effectiveness of her pretense-beliefs where there are personal gains in sight. My response to this is threefold.

First, it seems to me that there is no contradiction in simultaneously holding a genuine occurrent belief that nothing is morally anything (because error theory is true) and pretending to hold a genuine occurrent belief that some action is morally wrong. This seems to suggest that it is at least possible for the fictionalizer to attend to her belief that error theory is true and still pretend to hold moral beliefs and pretend to assert moral propositions. It at least cannot be taken for granted that attending to one’s belief that there are no dragons when role-playing an encounter with a dragon, or attending to one’s belief that error theory is true when role-playing oneself as moralizer necessarily stops the game or makes one step out of the fiction.

Second, a role-player engaged in a LARP can be aware that he is engaged with a fiction and still be *motivated* by this engagement. I claimed above that when a role-player – Stephen, for example – immerses in a character in a LARP, he focuses his attention on his pretense-feelings, pretense-beliefs, and pretense-desires, and less on his genuine feelings, beliefs and desires. But this does not mean that Stephen needs to forget or entirely suppress his genuine feelings, beliefs and desires, for I also claimed that it was possible for him to immerse in his character, and for his engagement with the fiction to have bearing on his motivation and behavior, without his holding a genuine belief with respect to any of the propositions that make up the fiction. When role-playing a sword-fight, Stephen can be aware, at least at some level, that he is handling a boffer sword. And it seems to me that if he is aware of the fact that he is handling a

¹⁴ Moreover, given that the attitudes of pretending to believe that *p* and genuinely believing that *not-p* are consistent – like fictionalism claims them to be – one might wonder what is meant by Oddie and Demetriou’s claim that “make-belief will [...] give way to genuine belief (Oddie and Demetriou 2010:201)” when push comes to practical shove.

boffer sword, he is also aware of the fact that he is engaged with a fiction. But as I argued above, this awareness need not hinder Stephen's pretense-beliefs (that he is fighting his antagonist using a real sword) and pretense-desires (to seriously injure or kill his antagonist) from having bearing on his motivation and action – even when Stephen is aware that he is engaged with a fiction, part of his behavior is most straightforwardly explained with reference to his engagement with this fiction.

Moreover, a role-player engaged in a LARP can be aware that he is engaged with a fiction *without attending* to his belief that he is so engaged. It seems to me that it is quite generally possible to be aware that one is situated in some particular context without having an occurrent belief that one is so situated. When Miranda, who is an ordinary philosophy student, is at the library reading a paper on fictionalism, for example, she is fully focused on the text on the computer screen in front of her. Her attention is directed at the argument formulated in the paper (presumably with a special attention to the claims made in the paragraph she is currently reading). She is not attending to her belief that she is at the library (and not, for example, in his study at home). But although Miranda does not attend to this belief, she can *continuously be aware*, at least at some level, that she is at the library. The reason for this is that some of her behavior is most straightforwardly explained by the assumption that she *is* so aware. If, for example, Miranda always sits properly in her chair when studying at the library, but does not always do so when studying at home, her behavior at the library is most straightforwardly explained by the assumption that when Miranda is at the library, she is at least at some level aware of this fact.

In the above two passages, I have argued that a role-player can be aware that he is engaged with a fiction, and that this awareness need not undermine the causal connection between his engagement with the fiction and motivation and action. Given this, the suggestion that engagement in revised moral practice is modeled on role-playing in LARPS seems to entail that for a fictionalizer engaged in revised moral practice, motivation can be sustained with continuous awareness that it is guided by fiction. If this is the case, then, contrary to what some have suggested, the fictionalizer does not need to suppress or silence her beliefs that engagement in revised moral practice is engagement with a fiction, that error theory is true, etc. in order to make sure that the causal connection between her engagement with the fiction and action does not break down.

A third response to this *prima facie* problem, which applies most directly to Köhler and Ridge's worry that the fictionalizer will take steps to undermine the motivational effectiveness of her moral pretense-beliefs when there are personal gains in sight, is that those considerations that provide her with non-moral reasons to stick to the fiction and her pretense-beliefs when tempted to deflect (see above) also seem to provide her with reasons to avoid taking steps to undermine the motivational effectiveness of her pretense-beliefs, even when there are personal gains in sight. In other words, it seems that the fictionalizer

has non-moral reasons not to undermine the motivational effectiveness of her pretense-beliefs that are intact in contexts where there are personal gains in sight.

As we saw above, some have claimed that in order to avoid forming genuine moral beliefs, fictionalizers engaged in revised moral practice must regularly attend to their genuine (second-order) beliefs about morality. When combined with the claim that fictionalizers need to suppress these beliefs for their engagement with the fiction to have bearing on motivation and action, so the argument goes, this reveals a *practical tension* in fictionalism. According to Olson, it is unclear whether fictionalism recommends the fictionalizer engaged in revised moral practice to (i) occasionally attend to, or (ii) suppress her genuine beliefs about morality (Olson 2014:189-90). This tension arises for versions of fictionalism which recommend a moral compartmentalization. The above arguments relieve this tension by illustrating that fictionalism need not recommend the fictionalizer to suppress her genuine beliefs about morality. Because her engagement with the moral fiction can have bearing on her motivation and action even if she is aware that she is engaged with a fiction, that error theory is true, and so on, fictionalism need not recommend any compartmentalization.

But even given that what I suggest above is correct, it *can* still be true that the fictionalizer needs to attend to her genuine beliefs about morality on a regular basis in order to avoid forming genuine moral beliefs. It seems to me, however, that this does not mean that fictionalism must include the recommendation that the fictionalizer regularly attends to these beliefs. The reason for this is that it need not be of any great consequence if the fictionalizer temporarily forgets, or stops being aware *even at some level*, that she is engaged with a fiction. As long as her moralized thinking serves her well in her everyday life, and her knowledge that ordinary moral practice is engagement with a fiction is available to her when she reflects critically, it is not clear that there would be any harm in temporarily forgetting about these facts when engaging in revised moral practice. For analogy, it is not clear that there would be any harm in our falling into absolutist thinking about time in our intuitive everyday thinking, as long as it serves us well practically and we can access relevant relativism about time in the relevant theoretical contexts. (Note that this is relevantly different from propagandism, as there is no effort to conceal the truth. We think as if absolutism about time were true although we do not believe that it is.)

Allowing that it may not matter – from the point of view of fictionalism – if the fictionalizer temporarily forgets or stops being aware that she is engaged with a fiction can be taken to suggest that it does not matter whether the agent really accepts the content of her moral judgments or merely role-plays such acceptance. As mentioned in section 7.4, this might, in turn, be taken to speak in favor of a conservatism-fictionalism hybrid theory which allows the agent to

slide from one state to the other without great consequence. In accepting this possibility I do not intend to make a concession to conservatism, for it seems to me that fictionalism of the kind suggested here has (at least) two major benefits over conservatism: First, it is more stable than conservatism in that it does not make moral motivation depend on the agent's repeatedly and reliably allowing herself to be seduced into holding genuine moral beliefs. According to fictionalism, there is another mechanism – viz. role-playing – that accounts for motivation. Second, it does not recommend moral compartmentalization or two level thinking. Because there is no contradiction in holding both a genuine occurrent belief that nothing is morally anything (because error theory is true) and pretending to hold a genuine occurrent belief that some action is morally wrong, and because moral motivation is accounted for by role-playing, the error theorist who has implemented fictionalism need not worry about the motivation-impeding effects of awareness that error theory is true, and so need not reserve such awareness to “critical contexts” such as the philosophy seminar room or the court-room.

In this section, I have argued that a fictionalism which models engagement in revised moral practice on role-playing in LARPs has potential to handle the *prima facie* problems facing fictionalism. Both problems concern fictionalism's potential to give an account of moral motivation. Referring to Manne (2015), I suggested that if genuine moral judgments can motivate us to act, then in order for them to do so, we need to see ourselves as occupying a specific role, namely that of an active participant in moral practice. An agent engaged in revised moral practice (i.e., a fictionalizer) can relate to herself in this manner by role-playing herself as moralizer.

The first *prima facie* problem for fictionalism was that it is unclear whether engagement with the moral fiction can have bearing on motivation and action. I argued, first, that because a role-player's engagement with the fiction of a LARP can have bearing on his motivation and action without causing genuine beliefs in the propositions of the LARP, it is reasonable to suppose that the fictionalizer's engagement with the moral fiction can have bearing on her motivation and action without causing genuine moral beliefs. Second, I suggested that the fictionalizer has non-moral reasons not to abandon her pretense-beliefs, and these are intact in contexts where deflecting would serve her self-interest.

The second *prima facie* problem was that it is unclear whether motivation can be sustained with awareness that it is guided by fiction. I gave a threefold response to this problem. First, I suggested that there is no contradiction in simultaneously believing that nothing is morally anything, and pretending to believe that some action is morally wrong. Second, because role-players can be aware that they are engaged with a fiction and still be motivated by this engagement, it is reasonable to suppose that fictionalizers can be aware that they are engaged with a fiction and still be motivated by this engagement.

Third, I suggested that the fictionalizer has non-moral reasons not to take steps to undermine the motivational effectiveness of her pretense-beliefs that are intact in contexts where there are personal gains in sight. I also explained how the fictionalism presently under consideration relieves the practical tension faced by versions of fictionalism which recommend a moral compartmentalization.

8.5 Conclusion

The topic of this and the previous chapter has been what practical implications error theory has for ordinary (first-order) moral thinking and discourse. More specifically, the central question of these chapters has been what we ought to do if we find out that ERROR THEORY is true. There are four competing answers to this question: abolitionism, propagandism, conservatism and fictionalism.

In chapter 7, I argued that ordinary (first-order) moral practice seems to be overall non-morally valuable in that it provides some pretty serious benefits throughout ordinary life. For this reason, abolitionism seems to come at a cost. Tentatively ruling out abolitionism left us with three competing suggestions as to what we ought to do if we find out that ERROR THEORY is true – viz. propagandism, conservatism and fictionalism. I argued that propagandism does not provide a complete response to the relevant question, and that there seems to be no reason to prefer propagandism to competing strategies the implementation of which would also leave moral practice more or less as it was prior to the discovery that ERROR THEORY is true, like conservatism. This left us with two competing answers to the central question– viz. conservatism and fictionalism. Third and last, I suggested that there are several similarities between conservatism and fictionalism, and that there are not sufficient reasons for preferring conservatism over fictionalism.

In this chapter, I provided considerations given which we ought to implement a version of fictionalism if we find out that ERROR THEORY is true. I argued that we can preserve the benefits of moral practice by role-playing morality on the model of role-playing in LARPs if we find out that ERROR THEORY is true, and explained how fictionalism can incorporate this suggestion.

I began by giving an account of Joyce's fictionalism, according to which we ought to carry on with morality as a fiction if we find out that error theory is true, where doing so essentially involves revising ordinary moral practice in such a way that it no longer involves any genuine moral beliefs or assertions, but merely pretense-beliefs and pretense-assertions. Second, I described two *prima facie* problems for fictionalism, both of which concern fictionalism's potential to give an account of moral motivation. Third, I explained how a fictionalism that incorporates the suggestion that we revise engagement in moral practice specifically on the model of role-playing in a live action role-playing can handle the *prima facie* problems for fictionalism.

I conclude that if ordinary (first-order) moral thinking and discourse is overall non-morally valuable to us, we ought to implement a version of fictionalism according to which engagement in revised moral practice is modeled on role-playing in live action role-playing games if we find out that ERROR THEORY is true; if we find out that ERROR THEORY is true, we ought to role-play morality.

References

Alvarez, Maria. 2017. "Reasons for action: justification, motivation, explanation". In Zalta, Edward N. (ed.). 2017. url: <https://plato.stanford.edu/archives/win2017/entries/reasons-just-vs-expl/>.

Ayer, A. J. 1936. *Language, Truth and Logic*. New York: Dover Publications Inc.

Beaver, David I. and Geurts, Bart. 2014. "Presupposition". In Zalta, Edward N. (ed.). 2014. url: <http://plato.stanford.edu/archives/win2014/entries/presupposition/>.

Bedke, Matthew S. 2010. "Might all normativity be queer?". *Australasian Journal of Philosophy* 88 (1): 41-58.

Bergamaschi Ganapini, Marianna. 2016. "Why we can still believe the error theory". *International Journal of Philosophical Studies* 24 (4): 523-536.

Björnsson, Gunnar, Strandberg, Caj, Francén Olinder, Ragnar, Eriksson, John and Björklund, Fredrik (eds.). 2015. *Motivational Internalism*. New York: Oxford University Press. E-book.

Blackburn, Simon. 1984. *Spreading the Word*. Oxford: Oxford University Press.

Blackburn, Simon. 1985. "Errors and the phenomenology of value". In Ted Honderich (ed.). *Morality and Objectivity: A Tribute to J. L. Mackie*. London: Routledge & Kegan Paul, 2-22.

Blackburn, Simon. 1993. *Essays in Quasi-realism*. Oxford: Oxford University Press.

Bloomfield, Paul (ed.). 2007. *Morality and Self-Interest*. New York: Oxford University Press.

Bøckman, Petter and Hutchison, Ragnhild (eds). 2005. *Dissecting larp – Collected papers for Knutepunkt 2005*. Oslo: The Knutepunkt 2005 Conference. E-book.

Brady, Michael (ed.). 2011. *New Waves in Metaethics*. Basingstoke: Palgrave Macmillan.

- Broad, C. D. 1951. "Hägerström's account of sense of duty and certain allied experiences". *Philosophy* 26 (97): 99-113.
- Burgess, John P. 1983. "Why I am not a nominalist". *Notre Dame Journal of Formal Logic* 24 (1): 93-105.
- Campbell, Erik. 2014. "Breakdown of moral judgment". *Ethics* 124 (3): 447-480.
- Clark, Maudemarie and Dudrick, David. 2007. "Nietzsche and moral objectivity: the development of Nietzsche's metaethics". In Brian Leiter and Neil Sinhababu (eds.). *Nietzsche and Morality*. Oxford: Oxford University Press, 192-226.
- Cole, Peter and Morgan, Jerry L. (eds.). 1975. *Syntax and Semantics 3: Speech Acts*. New York: Academic Press.
- Copp, David (ed.). 2006. *The Oxford Handbook of Ethical Theory*. New York: Oxford University Press.
- Cowie, Christopher. 2014. "Why companions in guilt arguments won't work". *The Philosophical Quarterly* 64 (256): 407-422.
- Cuneo, Terence. 2007. *The Normative Web: An Argument for Moral Realism*. Oxford: Oxford University Press.
- Cuneo, Terence and Christy, Sean. 2011. "The myth of moral fictionalism". In Michael Brady (ed.). *New Waves in Metaethics*. Basingstoke: Palgrave Macmillan, 85-102.
- Dahllöf, Mats. 1999. *Språklig Betydelse – En introduktion till Semantik och Pragmatik*. Lund: Studentlitteratur.
- Daly, Chris and Liggins, David. 2010. "In defence of error theory". *Philosophical Studies* 149 (2): 209-230.
- Dellsén, Finnur. 2017. "Understanding without justification or belief". *Ratio* 30 (3): 239-254.
- Dreier, James. 2004. "Meta-ethics and the problem of creeping minimalism". *Philosophical Perspectives* 18 (1): 23-44.
- Dreier, James. 2006. "Moral relativism and moral nihilism". In David Copp (ed.). *The Oxford Handbook of Ethical Theory*. New York: Oxford University Press, 240-264.

- Donnis, Jesper, Gade, Morten and Thorup, Line (eds.). 2007. *Lifelike*. Copenhagen: Projektgruppen KP07, Landsforeningen for Levende Rollespil, The Knudepunkt 2007 conference. E-book.
- Dworkin, Ronald. 1996. "Objectivity and truth: you'd better believe it". *Philosophy and Public Affairs* 25 (2): 87-139.
- Eklund, Matti. 2017. "Fictionalism". In Zalta, Edward N. (ed.). 2017. url: <http://plato.stanford.edu/archives/win2017/entries/fictionalism/>.
- Evers, Daan. 2016. "Jonas Olson's evidence of moral error theory". *Journal of Moral Philosophy* 13 (4):403-418.
- Finlay, Stephen. 2006. "The reasons that matter". *Australasian Journal of Philosophy* 84 (1): 1-20.
- Finlay, Stephen. 2008. "The error in the error theory". *Australasian Journal of Philosophy* 86 (3): 347-369.
- Finlay, Stephen and Schroeder, Mark. 2017. "Reasons for action: internal vs. external". In Zalta, Edward N. (ed.). 2017. url: <http://plato.stanford.edu/archives/fall2017/entries/reasons-internal-external/>.
- Fletcher, Guy. 2009. "Uneasy companions". *Ratio* 22 (3): 359-368.
- Foot, Philippa. 1972. "Morality as a system of hypothetical imperatives". *The Philosophical Review* 81 (3): 305-316.
- Foot, Philippa. 2002. *Virtues and Vices and Other Essays in Moral Philosophy*. Oxford: Clarendon.
- Garner, Richard. 1990. "On the genuine queerness of moral properties and facts". *Australasian Journal of Philosophy* 68 (2): 137-146.
- Garner, Richard. 1993. "Are convenient fictions harmful to your health?". *Philosophy East and West* 43 (1): 87-106.
- Garner, Richard. 2010. "Abolishing morality". In Richard Joyce and Simon Kirchin (eds.). *A World Without Values: Essays on John Mackie's Moral Error Theory*. Dordrecht: Springer Science+Business Media B.V., 217-233. E-book.
- Geuss, Raymond. 1997. "Nietzsche and morality". *European Journal of Philosophy* 5 (1): 1-20.

- Green, Mitchell. 2017. "Speech acts". In Zalta, Edward N. (ed.) 2017. url: <http://plato.stanford.edu/archives/win2017/entries/speech-acts/>.
- Grice, H. Paul. 1975. "Logic and conversation". In Peter Cole and Jerry L. Morgan (eds.). *Syntax and Semantics 3: Speech Acts*. New York: Academic Press, 41-58.
- Hägerström, Axel. 1953. *Inquiries Into the Nature of Law and Morals*. Uppsala: Almqvist & Wiksells Boktryckeri AB. Translated to English by C. D. Broad.
- Hägerström, Axel. 1964. *Philosophy and Religion*. London: George Allen & Unwin LTD. Translated to English by R. Sandin.
- Harding, Tobias. 2007. "Immersion revisited: role-playing as interpretation and narrative". In Jesper Donnis, Morten Gade and Line Thorup (eds.). *Lifelike*. Copenhagen: Projektgruppen KPO7, Landsforeningen for Levende Rollespil, The Knudepunkt 2007 conference, 25-33. E-book.
- Hedenius, Ingemar. 1940. "Om Hägerströms filosofi". *Tiden* 32 (1): 35-44.
- Hinckfuss, Ian. 1987. *The Moral Society: Its Structure and Effects*. Canberra: Department of Philosophy, Research School of Social Sciences, Australian National University.
- Holter, Matthijs. 2007. "Stop saying 'immersion'!". In Jesper Donnis, Morten Gade and Line Thorup (eds.). *Lifelike*. Copenhagen: Projektgruppen KPO7, Landsforeningen for Levende Rollespil, The Knudepunkt 2007 conference, 19-22. E-book.
- Honderich, Ted (ed.). 1985. *Morality and Objectivity: A Tribute to J. L. Mackie*. London: Routledge & Kegan Paul.
- Huemer, Michael. 2005. *Ethical Intuitionism*. Basingstoke: Palgrave Macmillan.
- Hussain, Nadeem J. Z. 2007. "Honest illusion: valuing for Nietzsche's free spirits". In Brian Leiter and Neil Sinhababu (eds.). *Nietzsche and Morality*. Oxford: Oxford University Press, 157-191.
- Hussain, Nadeem J. Z. 2010. "Error theory and fictionalism". In John Skorupski (ed.). *The Routledge Companion to Ethics*. London: Routledge, 335-345.
- Joyce, Richard. 2001. *The Myth of Morality*. Cambridge: Cambridge University Press.
- Joyce, Richard. 2005. "Moral fictionalism". In Mark Eli Kalderon (ed.). *Fictionalism in Metaphysics*. Oxford: Clarendon Press, 287-313.

- Joyce, Richard. 2006. *The Evolution of Morality*. Cambridge: The MIT Press.
- Joyce, Richard. 2007. "Morality, schmorality". In Paul Bloomfield (ed.). *Morality and Self-Interest*. New York: Oxford University Press, 51-76.
- Joyce, Richard. 2011a. "The error in 'The error in the error theory'". *Australasian Journal of Philosophy* 89 (3): 519-534.
- Joyce, Richard. 2011b. "The accidental error theorist". In Russ Shafer-Landau (ed.). *Oxford Studies in Metaethics. Volume 6*. Oxford: Oxford University Press, 153-180.
- Joyce, Richard. 2013. "Error theory". In Hugh LaFollette (ed.). *The International Encyclopedia of Ethics*. Oxford: Blackwell Publishing Ltd. E-book.
- Joyce, Richard. 2016a. "Moral anti-realism". In Zalta, Edward N. (ed.). 2016. url: <http://plato.stanford.edu/archives/win2016/entries/moral-anti-realism/>.
- Joyce, Richard. 2016b. "Reply to 'On the validity of a simple argument for moral error theory'". *International Journal of Philosophical Studies* 24 (4): 518-522.
- Joyce, Richard and Kirchin, Simon (eds.). 2010. *A World Without Values: Essays on John Mackie's Moral Error Theory*. Dordrecht: Springer Science+Business Media B.V. E-book.
- Joyce, Richard and Kirchin, Simon. 2010. "Introduction". In Richard Joyce and Simon Kirchin (eds.). *A World Without Values: Essays on John Mackie's Moral Error Theory*. Dordrecht: Springer Science+Business Media B.V., ix-xxiv. E-book.
- Kalderon, Mark Eli (ed.). 2005a. *Fictionalism in Metaphysics*. Oxford: Clarendon Press.
- Kalderon, Mark Eli. 2005b. *Moral Fictionalism*. Oxford: Oxford University Press.
- Kalf, Wouter Floris. 2013. "Moral error theory, entailment and presupposition". *Ethical Theory and Moral Practice* 16 (5): 923-937.
- Kant, Immanuel. 1996. *Practical Philosophy*. Cambridge: Cambridge University Press. Edited by Mary J. Gregor. Translated to English by Mary J. Gregor.
- Kim, John H. 2004. "Immersive story: a view of role-played drama". In Markus Montola and Jarkko Stenros (eds.). *Beyond Role and Play – Tools, Toys and Theory for Harnessing the Imagination*. Helsinki: The Solmukohta 2004 conference, Ropecon ry, 31-38. E-book.

- Köhler, Sebastian and Ridge, Michael. 2013. "Revolutionary expressivism". *Ratio* 26 (4): 428-449.
- Kornblith, Hilary. 1993. "Epistemic normativity". *Synthese* 94 (3): 357-376.
- LaFollette, Hugh (ed.). 2013, 2017. *The International Encyclopedia of Ethics*. Chichester: Wiley-Blackwell. E-book.
- Leiter, Brian. 1995. "Morality in the pejorative sense: on the logic of Nietzsche's critique of morality". *British Journal for the History of Philosophy* 3 (1): 113-145.
- Leiter, Brian. 1997. "Nietzsche and the morality critics". *Ethics* 107 (2): 250-285.
- Leiter, Brian and Sinhababu, Neil (eds.). 2007. *Nietzsche and Morality*. Oxford: Oxford University Press.
- Lewis, David. 1978. "Truth in fiction". *American Philosophical Quarterly* 15 (1): 37-46.
- Lillehammer, Hallvard. 2017. "Companions in guilt strategies". In Hugh LaFollette (ed.). *The International Encyclopedia of Ethics*. Oxford: Blackwell Publishing Ltd. E-book.
- Lillehammer, Hallvard and Möller, Niklas. 2015. "We can believe the error theory", *Ethical Theory and Moral Practice* 18 (3): 453-459.
- Lutz, Matt and Finlay, Stephen. 2015. "Review of 'Moral error theory: history, critique, defence' by Olson, Jonas". *Ethics* 125 (4): 1219-1225.
- Mackie, John Leslie. 1946. "A refutation of morals", *Australasian Journal of Psychology and Philosophy* 24 (1-2): 77-90.
- Mackie, John Leslie. 1977. *Ethics: Inventing Right and Wrong*. London: Penguin.
- Mackie, John Leslie. 1980. *Hume's Moral Theory*. London: Routledge.
- Mackie, John Leslie. 1982. *The Miracle of Theism*. Oxford: Oxford University Press.
- Mackie, John Leslie. 1985. *Persons and Values: Selected Papers Volume II*. Oxford: Oxford University Press.
- Manne, Kate. 2015. "Tempered internalism and the participatory stance". In Gunnar Björnsson, Caj Strandberg, Ragnar Francén Olinder, John Eriksson and Fredrik Björklund (eds.). *Motivational Internalism*. New York: Oxford University Press, 260-281. E-book.

- McPherson, Tristram and Plunkett, David (eds.). 2017. *The Routledge Handbook of Metaethics*, New York: Routledge. E-book.
- Miller, Alexander. 2013. *Contemporary Metaethics: An Introduction*. 2nd ed. Cambridge: Polity.
- Montola, Markus and Stenros, Jakko (eds.). 2004. *Beyond Role and Play – Tools, Toys and Theory for Harnessing the Imagination*. Helsinki: The Solmukohta 2004 conference, Ropecon ry. E-book.
- Munitz, Milton K. and Unger, Peter K. (eds.). 1974. *Semantics and Philosophy*. New York: New York University Press.
- Nietzsche, Friedrich. 2005. *The Anti-Christ, Ecce Homo, Twilight of the Idols And Other Writings*. Cambridge: Cambridge University Press. Edited by Aaron Ridley and Judith Norman. Translated to English by Judith Norman.
- Nietzsche, Friedrich. 2006a. *On the Genealogy of Morality*. Cambridge: Cambridge University Press. Edited by Keith Ansell-Pearson. Translated to English by Carol Diethe.
- Nietzsche, Friedrich. 2006b. *Daybreak: Thoughts on the Prejudices of Morality*. Cambridge: Cambridge University Press. Edited by Mudemarie Clark and Brian Leiter. Translated to English by R. J. Hollingdale.
- Nolan, Daniel, Restall, Greg and West, Caroline. 2005. "Moral fictionalism versus the rest". *Australasian Journal of Philosophy* 83 (3): 307-330.
- Oddie, Graham and Demetriou, Daniel. 2010. "The fictionalist's attitude problem". In Richard Joyce and Simon Kirchin (eds.). *A World Without Values: Essays on John Mackie's Moral Error Theory*. Dordrecht: Springer Science+Business Media B.V., 199-215. E-book.
- Olson, Jonas. 2011a. "In defense of moral error theory". In Michael Brady (ed.). *New Waves in Metaethics*. Basingstoke: Palgrave Macmillan, 62-84.
- Olson, Jonas. 2011b. "Getting real about moral fictionalism". In Russ Shafer-Landau (ed.). *Oxford Studies in Metaethics, Vol. 6*. Oxford: Oxford University Press, 181-204.
- Olson, Jonas. 2013. "Metaethics". In Hugh LaFollette (ed.). *The International Encyclopedia of Ethics*. Oxford: Blackwell Publishing Ltd. E-book.
- Olson, Jonas. 2014. *Moral Error Theory: History, Critique, Defence*. Oxford: Oxford University Press.

Olson, Jonas. 2016a. "Précis of moral error theory: history, critique, defence". *Journal of Moral Philosophy* 13 (4): 397-402.

Olson, Jonas. 2016b. "On the defensibility and believability of moral error theory". *Journal of Moral Philosophy* 13 (4): 461-473.

Olson, Jonas. 2017. "Error theory in metaethics". In Tristram McPherson and David Plunkett (eds.). *The Routledge Handbook of Metaethics*. New York: Routledge, 58-71. E-book.

Pagin, Peter. 2016. "Assertion". In Zalta, Edward N. (ed.). 2016. url: <http://plato.stanford.edu/archives/win2016/entries/assertion/>.

Petersson, Bo. 1973. *Axel Hägerströms värdeteori*. Ph.D. diss., Uppsala University.

Petersson, Bo. 2011. "Axel Hägerström and his early version of error theory". *Theoria* 77 (1): 55-70.

Pigden, Charles R. 2010. "Nihilism, Nietzsche, and the doppelganger problem". In Richard Joyce and Simon Kirchin (eds.). *A World Without Values: Essays on John Mackie's Moral Error Theory*. Dordrecht: Springer Science+Business Media B.V., 17-34. E-book.

Plato. 1997. *Plato: Complete Works*. Indianapolis: Hackett. Edited by John M. Cooper.

Ridge, Michael. 2013. "Non-Cognitivism". In Hugh LaFollette (ed.). *The International Encyclopedia of Ethics*. Oxford: Blackwell Publishing Ltd. E-book.

Rosati, Connie S. 2016. "Moral motivation". In Zalta, Edward N. (ed.). 2016. url: <http://plato.stanford.edu/archives/win2016/entries/moral-motivation/>.

Rowland, Richard. 2013. "Moral error theory and the argument from epistemic reasons". *Journal of Ethics & Social Philosophy* 7 (1): 1-24.

Sayre-McCord, Geoffrey (ed.). 1988. *Essays on Moral Realism*. Ithaca: Cornell University Press.

Sayre-McCord, Geoffrey. 1988. "Introduction: the many moral realisms". In Geoffrey Sayre-McCord (ed.). *Essays on Moral Realism*. Ithaca: Cornell University Press, 1-23.

Schacht, Richard. 1983. *Nietzsche*. London: Routledge & Kegan Paul. E-book.

- Schwitzgebel, Erik. 2015. "Belief". In Zalta, Edward N. (ed.). 2015. url: <http://plato.stanford.edu/archives/sum2015/entries/belief/>.
- Searle, John R. 1975. "The logical status of fictional discourse". *New Literary History* 6 (2): 319-332.
- Shafer-Landau, Russ (ed.). 2011. *Oxford Studies in Metaethics. Volume 6*. Oxford: Oxford University Press.
- Shepski, Lee. 2008. "The vanishing argument from queerness". *Australasian Journal of Philosophy* 86 (3): 371-387.
- Sinnott-Armstrong, Walter. 2006. *Moral Skepticisms*. New York: Oxford University Press.
- Skorupski, John (ed.). 2010. *The Routledge Companion to Ethics*. London: Routledge.
- Stalnaker, Robert C. 1974. "Pragmatic presuppositions". In Milton K. Munitz and Peter K. Unger (eds.). *Semantics and Philosophy*. New York: New York University Press, 197-213.
- Stevenson, C. L. 1937. "The emotive meaning of ethical terms". *Mind* 46 (181): 14-31.
- Streumer, Bart. 2013. "Can we believe the error theory?". *The Journal of Philosophy* 110 (4): 194-212.
- Streumer, Bart. 2016a. "No we cannot". *International Journal of Philosophical Studies* 24 (4): 537-546.
- Streumer, Bart. 2016b. "Why Jonas Olson cannot believe the error theory either". *Journal of Moral Philosophy* 13 (4): 419-436.
- Svavarsdóttir, Sigrun. 1999. "Moral cognitivism and motivation". *The Philosophical Review* 108 (2): 161-219.
- Toppinen, Teemu. 2016. "Is irreducible normativity impossibly queer?". *Journal of Moral Philosophy* 13 (4): 437-460.
- Utne, Torstein. 2005. "Live action role playing - teaching through gaming". In Petter Bøckman and Ragnhild Hutchison (eds.). *Dissecting larp – Collected papers for Knutepunkt 2005*. Oslo: The Knutepunkt 2005 Conference, 23-30. E-book.

van Roojen, Mark. 2016. "Moral cognitivism vs. non-cognitivism". In Zalta, Edward N. (ed.). 2016. url: <http://plato.stanford.edu/archives/win2016/entries/moral-cognitivism/>.

Williams, Bernard. 1981. *Moral Luck: Philosophical Papers 1973-1980*. Cambridge: Cambridge University Press.

Zalta, Edward N. (ed.). 2014-2017. *The Stanford Encyclopedia of Philosophy*, Stanford: Center for the Study of Language and Information, Stanford University. url: <http://plato.stanford.edu/>.

Sammanfattning

I våra vardagliga liv ställs vi ofta inför moraliska frågor. Exempelvis: Hur ska jag och min sambo fördela hushållssysslorna så att det blir rättvist? Är dödsstraff moraliskt försvarbart? Om alla barn har rätt till en trygg uppväxt, innebär det att flera än barnens föräldrar har en plikt att ge dem det?

När vi tänker på eller samtalar om denna typ av frågor, så tänker och talar vi som om det fanns *moraliska egenskaper*. Vi tänker och talar om handlingar och beteenden som *moraliskt rätta* eller *felaktiga*, *påbjudna* eller *förbjudna*. Vi tänker och talar om personer som *moraliskt goda* eller *onda*, *ärliga* eller *oärliga*, om politiska beslut och förslag som *rättvisa* eller *orättvisa*, *rättfärdiga* eller *orättfärdiga*. När vi tänker eller talar om någon eller något som moraliskt rätt, felaktigt, gott, ont, rättvist, orättvist (och så vidare), så tänker eller talar vi om personen eller saken som om den har en moralisk egenskap.

Våra moraliska övertygelser och omdömen spelar en viktig roll i våra liv. De påverkar våra tankar om vilken person vi vill vara eller bli, och hur vi relaterar till våra familjer, vänner, och andra som vi bryr oss om. De påverkar också vår syn på hur, och av vem, samhället skall styras, och vår syn på vad det är att vara en god partner, vän, och förälder. Exempelvis: Vi vill att hushållssysslorna ska fördelas rättvist. Vi känner att vi har ett ansvar gentemot miljön – och kanske också gentemot framtida generationer – att sortera våra sopor. Vi vill vara bra föräldrar och förebilder för våra barn, och försöker lära dem att säga ifrån när någon blir mobbad i skolan.

Denna avhandling handlar om *moralisk misstagsteori*. Misstagsteorin är en *metaetisk* teori. Metaetiken skiljer sig från den vardagliga etiken ("första ordningens" moraliskt tänkande och samtalande) på följande vis: När vi tänker på eller samtalar om etiska frågor i vardagliga sammanhang, så är vårt mål ofta att komma fram till svar på dessa etiska frågor. Vi vill exempelvis ta reda på huruvida dödsstraff, abort eller köttätande är moraliskt försvarbart, huruvida djur kan ha moraliska rättigheter, eller huruvida naturen har ett egenvärde. Inom metaetiken är målet snarare att komma fram till svar på frågor som har att göra med hur den vardagliga etiken kan beskrivas och analyseras. Metaetikern vill exempelvis ta reda på huruvida moraliska omdömen kan vara sanna (och i kraft av vad de i sådana fall är sanna), huruvida det är möjligt att ha moralisk kunskap (och hur vi i sådana fall kan nå sådan kunskap), och

huruvida det finns moraliska fakta (och om dessa fakta i sådana fall är objektiva eller subjektiva till sin natur).

De som försvarar misstagsteorin är en slags skeptiker. Om de har rätt, så är världen inte som vi tror att den är, och centrala antaganden som ingår i vår förståelse av oss själva och vår värld är felaktiga.

Misstagsteoretikerna menar att moralen oundvikligen innefattar ett grundläggande och allvarligt *misstag*. Misstaget består i att *det vi tänker* och *det vi hävdar* när vi tänker och samtalar om saker som om de har moraliska egenskaper aldrig är sant. Misstagsteoretikerna menar alltså att inga moraliska omdömen eller påståenden är sanna. Förklaringen till att det är så, menar många misstagsteoretiker, är att det helt enkelt inte finns några moraliska egenskaper eller fakta. Många argument för misstagsteorin är därför (ontologiska eller metafysiska) argument för tesen att det inte finns, eller *kan finnas*, sådana egenskaper eller fakta.

Förklaringen till att misstagsteoretiker ofta går till väga på detta sätt för att försvara tesen att inga moraliska omdömen eller påståenden är sanna är att de gör två grundläggande antaganden. Det första antagandet är att *funktionen* hos moraliska omdömen och påståenden är att beskriva världen. När vi fäller ett moraliskt omdöme eller gör ett moraliskt påstående, så tillskriver vi moraliska egenskaper till handlingar, personer, etc. Det andra är att detta innebär att ett moraliskt omdöme eller påstående kan vara *sant* endast om det finns (eller kan finnas) moraliska fakta eller egenskaper. Om det inte finns (eller kan finnas) moraliska fakta eller egenskaper, så är det oundvikligt att varje försök att beskriva sådana fakta, eller att tillskriva sådana egenskaper, kommer att misslyckas.

I den samtida filosofiska debatten är det många som har försvarat misstagsteorin genom att formulera argument som har att göra med vilka antaganden som görs, eller vad som tas för givet, i moraliskt tänkande och samtalande. Många sådana argument har fokuserat på antaganden som har att göra med vilken typ av *skäl för handlingar och beteenden* som moraliska egenskaper skulle ge oss om de fanns. En sådan misstagsteoretiker skulle kunna resonera så här: Om det fanns moraliskt rätta eller felaktiga handlingar, så skulle det åtminstone ibland vara fallet att någon hade *moraliska skäl* att agera eller bete sig på ett sätt snarare än ett annat. Men förslaget att någon har, eller kan ha, sådana skäl kan inte försvaras. Alltså finns inga moraliskt rätta eller felaktiga handlingar.

Det är förstås möjligt att komma fram till slutsatsen att misstagsteorin är sann med hjälp av andra typer av argument. Vissa kanske attraheras av misstagsteorin genom att reflektera över faktumet att det verkar möjligt att ge en fullständig naturvetenskaplig beskrivning av världen utan att nämna moraliska egenskaper. En sådan misstagsteoretiker skulle kunna argumentera så här: Om det fanns moraliska egenskaper, så skulle vi kunna identifiera något fenomen i världen som endast kunde förklaras med hänvisning till sådana

egenskaper. Men vi har inte lyckats identifiera något sådant fenomen. Alltså saknar vi skäl att tro att det finns moraliska egenskaper. Andra kanske attraheras av misstagsteorin efter att ha kommit fram till att ingen annan metaetisk teori håller för en kritisk granskning. En sådan misstagsteoretiker skulle kunna argumentera så här: Antingen är misstagsteorin riktig, eller så är någon annan metaetisk teori riktig (t.ex. realismen eller expressivismen). Men ingen annan metaetisk teori är riktig. Alltså är misstagsteorin riktig.

Denna avhandling har sju kapitel. I kapitel 1 introducerar jag misstagsteorin och förklarar, som jag har gjort ovan, vad avhandlingen handlar om. I kapitel 2 förklarar jag närmare vad som särskiljer misstagsteorier från andra uppfattningar: att de säger att inga ("positiva") moraliska omdömen eller påståenden är sanna. I kapitel 3 och 4 diskuterar jag sedan två argument för teorin. Båda argumenten har att göra med vad ett moraliskt skäl är (om det finns sådana skäl) och vad det skulle innebära för något att ha ett sådant skäl att agera eller bete sig på ett visst sätt snarare än ett annat. Jag menar att inget av argumenten ger oss starka skäl att tro på misstagsteorin. I kapitel 5 och 6 diskuterar jag två invändningar mot misstagsteorin. Båda invändningarna har att göra med vilka effekter (om några) misstagsteorin har på det kunskapsteoretiska området, exempelvis för frågor om vad det innebär att något är ett bevis eller ett gott skäl att hålla något för sant. Jag menar att ingen av invändningarna ger oss starka skäl att tro att misstagsteorin är felaktig. I kapitel 7 diskuterar jag vad som skulle hända med vardagligt ("första ordningens") moraliskt tänkande och samtalande om vi skulle upptäcka att misstagsteorin är sann. Jag menar att om sådant tänkande och samtalande är värdefullt för oss, så gör vi bäst i att anamma fikcionalismen om vi upptäcker att misstagsteorin är sann. Mer specifikt föreslår jag, i kapitel 7, att vi bör ändra vårt sätt att tänka och samtala moraliskt så att sådant tänkande och samtalande liknar hur man tänker och samtalar när man roll-spelar i ett levande rollspel ("lajv").

UMEÅ STUDIES IN PHILOSOPHY

Department of Historical, Philosophical and Religious Studies
Umeå University, Sweden

The series UMEÅ STUDIES IN PHILOSOPHY consists mainly of research communications (monographs, dissertations and collections of essays) in philosophy from the Department of Historical, Philosophical and Religious Studies at Umeå University. The series is not distributed through the bookstores, but individual volumes can be obtained from the department. Orders should be sent to Umeå Studies in Philosophy, Department of Historical, Philosophical and Religious Studies, Umeå University, SE-901 87 Umeå, Sweden.

1. JONAS NILSSON: *Rationality in Inquiry: On the Revisability of Cognitive Standards*, 2000.
2. PER NILSSON: *Naturen, vetenskapen & förnuftet: Upplysningens dialektik och det andra moderna*, 2001.
3. JAYNE M. WATERWORTH: *Living in the Light of Hope: An Investigation into Agency and Meaning*, 2001.
4. ANDERS ODENSTEDT: *Cognition and Cultural Context: An Inquiry Into Gadamer's Theory of Context-Dependence*, 2001.
5. RÖGNVALDUR INGTHORSSON: *Time, Persistence, and Causality: Towards a Dynamic View of Temporal Reality*, 2002.
6. BENGT LILIEQUIST: *Ludwik Flecks jämförande kunskapsteori*, 2003.
7. PETER NILSSON: *Empathy and Emotions: On the Notion of Empathy as Emotional Sharing*, 2003.
8. ANDERS BERGLUND: *From Conceivability to Possibility: An Essay in Modal Epistemology*, 2005.
9. LARS SAMUELSSON: *The Moral Status of Nature: Reasons to Care for the Natural World*, 2008.
10. EBBA GULLBERG: *Objects and Objectivity: Alternatives to Mathematical Realism*, 2011.

11. JESPER ÖSTMAN: *It's All in the Brain: A Theory of the Qualities of Perception* 2013.
12. EMMA BECKMAN: *Mistaken Morality? An Essay on Moral Error Theory*, 2018.